

Правительство Российской Федерации  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ  
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»  
(НИУ ВШЭ)

УДК 33:005.311.6

Рег. № НИОКТР АААА-А20-120070290164-6

Рег. № ИКРБС

УТВЕРЖДАЮ  
Проректор НИУ ВШЭ  
канд. экон. наук  
\_\_\_\_\_ М.М. Юдкевич  
«\_\_» \_\_\_\_\_ 2020 г.

ОТЧЕТ  
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

ПРИНЯТИЯ РЕШЕНИЙ И АНАЛИЗ ДАННЫХ В СОЦИАЛЬНО-  
ЭКОНОМИЧЕСКИХ И ПОЛИТИЧЕСКИХ СИСТЕМАХ  
(заключительный)

Руководители НИР:

Директор международного центра  
анализа и выбора решений,  
д-р техн. наук, проф.

\_\_\_\_\_ Ф.Т. Алескеров

Главный научный сотрудник  
международного центра  
анализа и выбора решений,  
PhD

\_\_\_\_\_ Э.С. Маскин

Москва 2020

## СПИСОК ИСПОЛНИТЕЛЕЙ

### Руководители НИР:

Директор международного центра анализа и выбора решений, д-р техн.наук, проф.	_____	Ф.Т. Алескеров (введение, заключение)
	подпись, дата	
Главный научный сотрудник международного центра анализа и выбора решений, PhD	_____	Э. Маскин (пункт 4.1.1)
	подпись, дата	

### Исполнители:

Заместитель директора центра, д-р физ.- мат. наук, проф.	_____	А.Е. Лепский (подраздел 2.6)
	подпись, дата	
Ведущий эксперт, канд. техн. Наук	_____	В.А. Саркисянц (аннотации)
	подпись, дата	
Эксперт, канд. техн.наук	_____	А.А. Рубчинский (подраздел 3.2)
	подпись, дата	
Ведущий научный сотрудник, д-р техн. наук, проф.	_____	А.С. Беленький (пункты 4.1.2, 4.2.2)
	подпись, дата	
Ведущий научный сотрудник, д-р физ.-мат. наук, проф.	_____	А.А. Лазарев (пункт 4.2.1)
	подпись, дата	
Ведущий научный сотрудник, д-р техн. наук, проф.	_____	Б.Г. Миркин (подразделы 3.1.2, 3.1.3)
	подпись, дата	
Ведущий научный сотрудник, д-р техн.наук, профессор-исследователь	_____	В.В. Подиновский (подраздел 1.1)
	подпись, дата	
Старший научный сотрудник, канд. экон. Наук	_____	Д.С. Карабекян (пункт 1.2.1)
	подпись, дата	
Старший научный сотрудник, канд. экон. Наук	_____	А.В. Карпов (подраздел 1.3)
	подпись, дата	

Старший научный сотрудник, канд. физ.-мат. наук	_____ подпись, дата	В.С. Молоствов (пункт 4.2.3)
Старший научный сотрудник, канд. техн.наук, доцент	_____ подпись, дата	А.Л. Мячин (подраздел 3.3)
Старший научный сотрудник, канд. экон. наук, доцент	_____ подпись, дата	Г.И. Пеникас (пункты 6.1.4, 6.1.5)
Старший научный сотрудник, канд. физ.-мат. наук	_____ подпись, дата	А.Н. Субочев (подраздел 1.4)
Старший научный сотрудник, канд. наук по прикл. мат.	_____ подпись, дата	С.В. Швыдун (подраздел 2.2)
Старший научный сотрудник, канд. техн. Наук	_____ подпись, дата	В.И. Якуба (пункт 1.2.4)
Научный сотрудник, PhD	_____ подпись, дата	Э. Доган (реферат)
Научный сотрудник, канд. физ.- мат. Наук	_____ подпись, дата	Л.Г. Егорова (подраздел 1.5)
Научный сотрудник, канд. наук по прикл. мат.	_____ подпись, дата	М.А. Марон (подраздел 2.7)
Младший научный сотрудник, канд.физ.-мат.наук	_____ подпись, дата	Ю.А. Веселова (пункт 1.2.3)
Стажер-исследователь	_____ подпись, дата	Е.А. Борсук (подраздел 5.2)
Стажер-исследователь, аспирант	_____ подпись, дата	Г.Т. Броницкий (список исп. источни- ков)
Стажер-исследователь, аспирант	_____ подпись, дата	С.С. Демин (подраздел 5.6)
Стажер-исследователь, аспирант	_____ подпись, дата	М.Д. Ермолова (пункты 6.1.2, 6.1.3)

Стажер-исследователь	_____	Е.С. Замятнина (подраздел 5.7)
	подпись, дата	
Стажер-исследователь	_____	М.М. Земцова (пункт 5.4.2)
	подпись, дата	
Стажер-исследователь	_____	А.А. Иванов (пункт 1.2.2)
	подпись, дата	
Стажер-исследователь	_____	М.С. Курапова (подраз- дел 2.1)
	подпись, дата	
Стажер-исследователь	_____	В. Кучински (пункт 5.5.2)
	подпись, дата	
Стажер-исследователь, студент	_____	И.-А.И. Латыпов (под- раздел 2.5)
	подпись, дата	
Стажер-исследователь, студент	_____	А.В. Лиханова (пункт 2.3.4)
	подпись, дата	
Стажер-исследователь, студент	_____	А.С. Ломакина (пункт 6.1.6)
	подпись, дата	
Стажер-исследователь, канд. наук по прикл. мат.	_____	Т.А. Ломоносов (полраздел 6.3)
	подпись, дата	
Стажер-исследователь, аспирант	_____	Н.Г. Мецерыкова (подраздел 2.4)
	подпись, дата	
Стажер-исследователь	_____	А.Н. Резяпова (пункт 2.3.2)
	подпись, дата	
Стажер-исследователь, студент	_____	А.А. Роман (пункты 2.3.1, 2.3.3)
	подпись, дата	
Стажер-исследователь, студент	_____	В.М. Семчукова (подраздел 5.3)
	подпись, дата	
Стажер-исследователь, студент	_____	Ю.Б. Скрупская (подраздел 5.1)
	подпись, дата	

Стажер-исследователь	_____	М.А. Сурков (пункт 6.1.1)
	подпись, дата	
Стажер-исследователь	_____	Д.Н. Тверской (подраздел 6.2)
	подпись, дата	
Стажер-исследователь, студент	_____	Д.С. Ткачев (пункт 5.4.1)
	подпись, дата	
Стажер-исследователь, канд. комп. Наук	_____	Д.С. Фролов (подразделы 3.1.1, 3.1.4)
	подпись, дата	
Стажер-исследователь, студент	_____	Д.А. Шадрин (пункт 5.5.1)
	подпись, дата	
Нормоконтроль	_____	А.А. Павлова
	подпись, дата	

## РЕФЕРАТ

Отчет 344 с., 1 кн., 57 рис., 39 табл., 312 источн., 2 прил.

### ТЕОРИЯ КОЛЛЕКТИВНОГО ВЫБОРА, ТЕОРИЯ ПРИНЯТИЯ РЕШЕНИЙ, МЕТОДЫ ОПТИМИЗАЦИИ, АНАЛИЗ ДАННЫХ, ТЕОРИЯ ФУНКЦИЙ ДОВЕРИЯ

Объектами исследования были модели манипулируемости правил агрегирования предпочтений и модели профилей предпочтений, модели агрегирования предпочтений в рамках теории коллективного выбора, модели взаимозависимости и группового влияния в сетевых структурах, модели оценивания схожести и агрегирования слоёв многослойной сетевой структуры, гравитационная модель торговли, модель подъема тематического кластера в таксономии предметной области, модель оценивания экономической целесообразности создания грузового транспортного узла, модель размещения зарядных станций автомобилей, модели справедливого распределения спорных территорий, модели оболочечного анализа данных, модель образования теневой банковской системы, модели специализации в многоклеточных системах.

Целью проекта были разработка и исследование моделей и методов принятия решений и анализа данных в социально-экономических системах.

Методологической базой проекта были: теория коллективного и индивидуального выбора, теория многокритериального принятия решений, кластерный и сетевой анализ данных, методы оптимизации, теория функций доверия, методы системного анализа, методы теории вероятностей и математической статистики, методы теории игр и др.

Область применения полученных результатов: региональные транспортные системы, инвестиционный анализ, банковское регулирование, сетевой анализ данных, теория коллективного выбора, многокритериальное принятие решений, миграционное регулирование, управление превентивными противопожарными мероприятиями в регионах, межгосударственное взаимодействие и др.

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	13
1 Исследование моделей выбора решений и анализа предпочтений ....	22
1.1 Разработка моделей средних значений выборок в рамках теории важности критериев .....	22
1.1.1 Введение.....	22
1.1.2 Сведения о средних величинах.....	23
1.1.3 Сведения из теории многокритериальной оптимизации .....	26
1.1.4 Определение средних величин как недоминируемых точек	31
1.1.5 Средние величины при равноважных критериях .....	33
1.1.6 Средние величины при равноважных критериях и отсутствии компенсации удаленностей.....	35
1.1.7 Средние величины при равноважных критериях со шкалой первой порядковой метрики .....	36
1.1.8 О построении множеств средних величин .....	40
1.1.9 О сравнении множественных средних по величине.....	41
1.1.10 Случай данных с неопределенностями.....	42
1.1.11 Заключение .....	44
1.2 Разработка и анализ моделей расчета степени манипулируемости правил агрегирования предпочтений .....	45
1.2.1 Введение.....	45
1.2.2 Алгоритмы для оценки индексов манипулируемости процедур агрегирования.....	46
1.2.3 Безопасное коалиционное манипулирование .....	64
1.2.4 Заключение .....	69
1.3 Нахождение вероятностей исследование однопиковых предпочтений .....	70
1.3.1 Введение.....	70
1.3.2 Основные понятия.....	74

1.3.3 Вероятностные модели .....	75
1.4 Исследование турнирных способов выбора оптимальных вариантов.....	77
1.4.1 Описание модели.....	78
1.4.2 Характеристические теоремы .....	79
1.4.3 Проблема выбора из бесконечного множества.....	81
1.5 Анализа предпочтений студентов относительно форматов онлайн-обучения .....	83
1.5.1 Введение.....	83
1.5.2 Основная часть .....	84
1.5.3 Некоторые выводы.....	90
2 Исследование моделей сетевого анализа данных.....	92
2.1 Введение.....	92
2.2 Разработка обобщенного индекса LRIC, учитывающего влияние в сетевых структурах .....	100
2.2.1 Описание модели.....	100
2.2.2 Применение обобщенного индекса LRIC к анализу сети международной торговли.....	103
2.2.3 Применение индекса LRIC к анализу сети взаимодействий террористических групп.....	105
2.3 Разработка и применение индексов центральности, учитывающих индивидуальные и групповые влияния.....	108
2.3.1 Математическая модель новых индексов центральности .	108
2.3.2 Анализ сети международной миграции.....	109
2.3.3 Анализ сети общей торговли .....	113
2.3.4 Сетевой анализ экспорта и импорта нефти и газа .....	117
2.4 Оценивание схожести слоёв сети торговли продовольствием...	122
2.4.1 Описание модели схожести слоёв сети .....	122
2.4.2 Применение модели к сети торговли продовольствием ....	124



2.5 Исследование многосторонних торговых потоков в гравитационной модели торговли .....	126
2.6 Оценивание центральностей многослойной сети методами теории функций доверия .....	129
2.6.1 Постановка задачи.....	129
2.6.2 Основные сведения из теории функций доверия.....	130
2.6.3 Основные обозначения и классические центральности графа .....	132
2.6.4 Оценивание центральностей однослойной сети методами теории функций доверия .....	133
2.6.5 Многослойные сети и агрегирование центральностей .....	137
2.7 Некоторые выводы .....	141
3 Исследования по анализу текстов и кластерному анализу .....	144
3.1 Разработка гибридного метода экономного подъема нечеткого тематического кластера в таксономии .....	144
3.1.1 Постановка задачи.....	144
3.1.2 Экономный подъем нечеткого тематического кластера в таксономии: модель и метод .....	147
3.1.3 Изучение направлений исследований на основе коллекций научных публикаций с помощью таксономии предметной области .....	154
3.1.4 Некоторые выводы .....	167
3.2 Разработка и исследование метода альтернирующих дихотомий для кластеризации данных .....	170
3.2.1 Постановка задачи.....	170
3.2.2 Метод альтернирующих дихотомий .....	172
3.2.3 Применение МАД для анализа фондового рынка .....	177
3.2.4 Некоторые выводы .....	186
3.3 Исследование государственной состоятельности в странах мира методами кластерного анализа .....	187

3.3.1 Обзор литературы .....	187
3.3.2 Общая постановка задачи исследования .....	189
3.3.3 Оценивание государственной состоятельности методами анализа паттернов .....	191
3.3.4 Оценивание государственной состоятельности с помощью иерархической кластеризации .....	193
3.3.5 Сопоставление результатов порядково-инвариантной паттерн-кластеризации и иерархического кластерного анализа.....	195
3.3.6 Динамический анализа траекторий развития стран .....	196
3.3.7 Некоторые выводы.....	197
4 Исследования моделей по оптимальному размещению объектов и привлечению частных инвестиций .....	198
4.1 Теоретико-игровой подход к оценке экономической целесообразности создания нового грузового транспортного узла.....	198
4.1.1 Постановка основных задач исследования.....	198
4.1.2 Краткий обзор состояния работ в области исследования..	206
4.2 Исследование задачи оптимизации размещения в городских районах станций по зарядке и по замене батарей для питания электромобилей .....	211
4.2.1 Постановка основных задач исследования.....	211
4.2.2 Краткий обзор подходов к математическому моделированию проблем, близких к проблеме, рассматриваемой в настоящем исследовании.....	216
4.2.3 Построение математической модели рассматриваемой задачи .....	227
5 Анализ данных и принятие решений в задачах социально- экономического развития регионов .....	238
5.1 Миграция в арктических регионах.....	238

5.1.1 Общий анализ проблемы миграции в арктическом регионе .....	238
5.1.2 Анализ данных по миграции в арктическом регионе .....	240
5.2 Распределение спорных территорий арктического региона .....	247
5.2.1 Общий анализ проблемы спорных территорий арктического региона .....	247
5.2.2 Анализ методов справедливого распределения спорных территорий и их модификация .....	248
5.3 Анализ уязвимости социально-экономического развития Арктики из-за изменения климата .....	257
5.4 Оценка водообеспеченности стран Африки.....	260
5.4.1 Общий анализ проблемы водообеспеченности стран Африки .....	260
5.4.2 Анализ данных по водообеспечению стран Африки .....	261
5.5 Продовольственная безопасность ряда африканских стран и её зависимость от изменения климата .....	265
5.5.1 Общий анализ проблемы продовольственной безопасности африканских стран .....	265
5.5.2 Аналитический обзор научной литературы по исследованию продовольственной безопасности стран Африки .....	266
5.6 Применение интервального оболочечного анализа данных к оцениванию эффективности превентивных противопожарных мероприятий.....	268
5.6.1 Постановка задачи.....	268
5.6.2 Основная модель .....	271
5.7 Некоторые выводы .....	274
6 Исследование моделей регулирования, распространения и специализации .....	279

6.1 Исследование моделей регулирования в банковской сфере:	
подход на основе внутренних рейтингов для расчета требований к капиталу.....	279
6.1.1 Постановка задачи.....	279
6.1.2 Описание анализируемой модели .....	280
6.1.3 Включение режима ПВР и регулирования капитала в анализируемую модель.....	282
6.1.4 Вычет потерь вкладчиков из благосостояния .....	288
6.1.5 Сравнение прибыли в традиционном и теневом банкинге	289
6.1.6 Некоторые выводы.....	291
6.2 Модели специализации в абстрактных системах: случай структурных ограничений, задаваемых неравенствами .....	292
6.2.1 Постановка задачи.....	292
6.2.2 Основные результаты .....	295
6.2.3 Некоторые выводы.....	298
6.3 Разработка матричного метода линеаризации систем уравнений в частных производных на постоянных решениях.....	298
ЗАКЛЮЧЕНИЕ .....	300
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ .....	304
ПРИЛОЖЕНИЕ А Таблицы к подразделу 1.5 .....	334
ПРИЛОЖЕНИЕ Б Таблицы к разделу 5 .....	342

## ВВЕДЕНИЕ

В данном отчете приведены результаты работы в 2020 году сотрудников Международного центра анализа и выбора решений (МЦАВР) НИУ ВШЭ над проектом «Принятия решений и анализ данных в социально-экономических и политических системах».

В соответствии с техническим заданием основными научными направлениями исследований были:

- модели выбора решений и анализа предпочтений;
- модели сетевого анализа данных;
- анализ текстов и кластерный анализ данных;
- модели оптимального размещения объектов и привлечения частных инвестиций;
- анализ данных и принятие решений в задачах социально-экономического развития регионов;
- модели регулирования, распространения и специализации.

В настоящем проекте были продолжены исследования 2019 г. по выполнению НИР «Исследование моделей принятия решений и анализа сложно структурированных данных» [1]. В частности, в 2019 г. были получены следующие результаты:

- в рамках исследования манипулируемости правил агрегирования предпочтений были проведены вычислительные эксперименты для расчета вероятности индивидуального манипулирования при наличии контругроз для 11 правил коллективного выбора;
- исследованы различные подходы и методы анализа решений в условиях неопределенности, основанные на теории важности критериев и не предполагающие использования функций полезности и количественных вероятностей;
- выполнен аксиоматический анализ способов выбора оптимальных вариантов и построения рейтинга альтернатив;

- в рамках исследования профилей предпочтений получены: точная формула для числа группоразделимых профилей предпочтений; рекурсивная формула для числа нарцисстических группоразделимых профилей предпочтений;
- в рамках исследования стабильных паросочетаний разработано новое расширение условия последовательного предпочтения, которое приводит к единственному стабильному паросочетанию для любого подмножества агентов;
- разработан метод оценивания влияния в мультиплексных сетевых структурах; предложенная модель применена к сетям международной торговли продовольствием и торговли вооружением;
- предложен и исследован ряд мер оценивания важности рёбер в сетевых структурах;
- проведен сетевой анализ международной миграции с 1970 по 2014 гг. на основе данных по потокам миграции между странами; определены рейтинги влиятельных стран в процессе международной миграции с помощью индексов центральностей, как классических, так и индексов ближних и дальних взаимодействий;
- проведен сетевой анализ международной торговли с учетом прямого и непрямого взаимодействия между странами;
- в рамках исследования методов анализа текстов было введено понятие наиболее подходящего обобщения нечеткого множества тем, сформированного листьями корневого дерева таксономии; предложенный метод обобщения “поднимает” нечеткое множество тем к узлам таксономии более высокого ранга наиболее “тесным” способом; метод апробирован на текстовой коллекции из 18000 научных публикаций по тематике анализа данных, и таксономии Науки данных, сформированной на основе Таксономии Computing Machinery Computing Classification System 2012, ACM-CCS;

– в рамках разработки методов анализа паттернов предложен алгоритм выделения симметричных паттернов на больших и сильно зашумленных данных; показано, что симметрия паттернов сохраняется при любом порядке их формирования;

– в рамках исследования конкурентной кооперации в Арктике и на основе результатов учебного и исследовательского курсов в рамках проекта «Общество и передовые технологии в Арктике» (SATA) введена и описана концепция цикла SATA, которая отражает междисциплинарность применения фундаментальных исследований для улучшения условий жизни в Арктике с использованием передовых технологий;

– предложена математическая модель для оценки необходимого объема инвестиций в развитие региональной инфраструктуры грузоперевозок; предложенная модель апробирована на модельных примерах грузовых транспортных систем; сформулированы некоторые рекомендации по использованию этой модели;

– выявлены и описаны классы системных проблем, связанных с формированием государственно-частного партнерства (на примере энергетических систем);

– модифицированы математические модели, на основе которых переформулированы математические задачи оценки объема инвестиций, требуемого для размещения цепи зарядных станций электромобилей на шоссе; доказаны соответствующие математические утверждения;

– показана возможность применения методов теории неаддитивных мер и функций доверия к решению ряда задач политологического анализа: нахождения наиболее значимых для голосования групп вопросов, оценивания политической неоднородности общества, оценивания значимости позиций отдельных партий для голосования, оценивания конфликтности позиций партий по значимым вопросам; исследование выполнено по базе данных интернет-опросов партий в Германии перед выборами в Бундестаг в 2013г. и результатов этих выборов;

– проанализированы основные модели и методы оценивания поляризации и политического влияния, в том числе и модели, учитывающие предпочтения по формированию коалиций;

– в рамках анализа моделей в банковской сфере проанализировано влияние модельного риска на точность оценок величины риск-взвешенных активов, полученных с помощью подхода на основе внутренних рейтингов; предложены меры по учету такого влияния в задаче повышения сопоставимости оценок кредитного риска;

– в рамках исследования структурной сложности графов разработана новая модификация показателя сложности графа, основанная на построении семейства графовых дихотомий; с помощью предложенного подхода анализируются графы для данных фондового рынка S&P-500 за период 1991-2019 гг. с целью прогнозирования особых периодов в поведении фондового рынка;

– исследованы явные двухуровневые конечно-разностные схемы решения полной системы уравнений газовой динамики для случая произвольной фоновой скорости;

– была поставлена и исследована оптимизационная задача возникновения специализации в абстрактной многоклеточной системе на основе максимизации эффективности функционирования этой системы.

Некоторые исследования 2020 года являются продолжением исследований 2019 г. Это, например, исследования по разработке и анализу моделей расчета степени манипулируемости правил агрегирования предпочтений, исследования по теории важности критериев, исследования по анализу профилей предпочтений, исследования по сетевому анализу данных, по анализу текстов, по анализу структурной сложности графов и др.

Некоторые исследования этого года предполагается продолжить в следующем году.

При выполнении данного проекта использовались современные методы анализа данных и машинного обучения, теории принятия решений, методы оптимизации, теории графов, теории многокритериального и коллективного



выбора, теории вероятностей, теории функций доверия, методы исследования операций.

Ряд прикладных задач анализа данных и принятия решений был решен с использованием открытых баз данных. Решения многих задач были промоделированы с использованием прикладного программного обеспечения, а некоторые задачи потребовали разработки комплексов программ.

Отчет состоит из 6 разделов, соответствующих основным направлениям исследований.

В первом разделе представлены результаты по исследованию моделей выбора решений и анализа предпочтений.

В подразделе 1.1. приведены результаты по разработке и исследованию моделей средних значений выборок в рамках теории важности критериев, основанных на идеях многокритериальной оптимизации.

Подраздел 1.2. посвящен обзору результатов текущего года продолжающегося исследования по разработке и анализу моделей расчета степени манипулируемости правил агрегирования предпочтений. В частности, для моделей голосования Impartial Culture и Impartial Anonymous Culture, разных способов построения расширенных предпочтений, количества альтернатив  $3 \div 5$ , участников голосования  $3 \div 100$  произведены расчеты, показывающие, какая процедура агрегирования будет наименее манипулируемой. В исследовании безопасности коалиционного манипулирования найдены условия, когда манипулирование всегда безопасно и когда существует небезопасно манипулируемый профиль для двухступенчатой мажоритарной системы и для правила Коупленда.

В подразделе 1.3 приведены результаты нахождения вероятностей однопиковых профилей предпочтений для различных вероятностных моделей генерирования профилей.

В подразделе 1.4 найдены достаточные условия непустоты множества обобщенных турнирных решений, основанных на принципе устойчивости.

Кроме того, показано, что эти же условия будут справедливы для некоторых турнирных решений и в случае бесконечного множества вариантов.

Последний подраздел 1.5 данного раздела отчета, носит прикладной характер. В нем приведены результаты анализа предпочтений студентов относительно форматов онлайн-обучения. В частности, применяя методы анализа латентных классов и дисперсионного анализа к обработке мониторинговых данных, выделены и проанализированы различные группы студентов по их отношению к переводу занятий в онлайн формат.

Во втором разделе отчета приведены результаты исследования моделей сетевого анализа данных. Этот раздел состоит из введения (подраздел 2.1), в котором дан общий обзор работ Центра в области сетевого анализа данных, и ряда подразделов, в которых отражены результаты отдельных исследований.

Так в подразделе 2.2 приведены результаты по разработке обобщенного индекса дальних взаимодействий LRIC, учитывающего взаимозависимость между элементами сети. Эти индексы применены к анализу сетей международной торговли и взаимодействия террористических групп.

Подраздел 2.3 посвящен описанию результатов по разработке индексов центральности, учитывающих индивидуальные и групповые влияния вершин сети. Эти индексы применены к анализу сетей международной миграции и межстрановой торговли.

В подразделе 2.4 приведены результаты по разработке модели оценивания схожести слоёв многослойной динамической сетевой структуры. Эта модель применена к анализу сети международного экспорта/импорта продовольствия.

В подразделе 2.5 рассмотрена модификация гравитационной модели, включающей в себя индексы центральности сети. Эта модель применена к анализу многосторонних торговых потоков.

В подразделе 2.6 приведены предварительные результаты по разработке и исследованию способа оценивания центральностей многослойной сети методами теории функций доверия. Доказаны взаимосвязи полученных оценок с

агрегированием классических послойных центральностей в некоторых частных случаях. В подразделе 2.7 приведены некоторые выводы по второму разделу.

Третий раздел отчета посвящен описанию результатов исследований по анализу текстов и кластерному анализу данных. Эти исследования осуществлялись в направлении кластерного анализа в таксономии предметной области, разработке метода альтернирующих дихотомий для кластеризации данных, применения методов кластерного анализа к исследованию государственной состоятельности в странах мира.

В подразделе 3.1 описываются результаты разработки гибридной модели и метода экономного подъема нечеткого тематического кластера в таксономии предметной области. Этот метод применен к выявлению исследовательских тенденций по анализу статей, опубликованных в журналах издательств Springer и Elsevier по тематике Наука о Данных.

В подразделе 3.2 приведены результаты по разработке и исследованию метода альтернирующих дихотомий для кластеризации данных. Этот метод применен для анализа фондового рынка.

В подразделе 3.3 приведены результаты исследования государственной состоятельности в странах мира в период 1996-2015гг с помощью двух различных методов (иерархической кластеризации и порядково-инвариантной паттерн-кластеризации).

В четвертом разделе отчета приведены результаты исследования моделей по оптимальному размещению объектов и привлечению частных инвестиций.

Так подраздел 4.1 посвящен описанию теоретико-игрового подхода к оценке экономической целесообразности создания нового грузового транспортного узла. Это продолжающееся исследование. В этом разделе, в частности, сформулировано и доказано достаточное условие конкурентоспособности тарифов на услуги в таком узле. Продемонстрирована возможность формули-

ровки и решения задачи об оценке финансовой целесообразности строительства и эксплуатации нового транспортного узла, как задачи нахождения максимума максимина разности двух билинейных функций.

В подразделе 4.2 приведены результаты исследования задачи оптимизации размещения в городских районах станций по зарядке и по замене батарей для питания электромобилей.

В пятом разделе отчета отражены результаты по применению методов анализа данных к решению ряда прикладных задач социально-экономического развития регионов.

Так в подразделе 5.1 приведены результаты анализа изменения населения в Арктике из-за естественного прироста и чистой миграции населения.

В подразделе 5.2 проанализированы существующие методы распределения спорных территорий и предложен ряд новых модификаций этих методов. Продемонстрирована работа этих методов на примере раздела спорных территорий стран Арктического региона.

В подразделе 5.3 отражены результаты по анализу влияния изменения климата на социально-экономическое развитие населения Арктики.

В подразделах 5.4 и 5.5 проведен анализ климатических и географических данных стран Африки относительно влияния этих факторов на водообеспеченность, водопотребление и на продовольственную безопасность этих стран.

В подразделе 5.6 приведены результаты по разработке модификаций интервального оболочечного анализа данных. Эти методы были применены к оцениванию эффективности превентивных противопожарных мероприятий.

Наконец шестой раздел посвящен исследованию некоторых моделей регулирования, распространения и специализации.

Так, в подразделе 6.1 приведены результаты по анализу одной популярной модели образования теневой банковской системы. Эта модель модифицирована, и с помощью этой модификации показано, что политика регулятора не может основываться на поддержке банковского теневого сектора.

Подраздел 6.2 содержит результаты исследования моделей специализации в абстрактных системах. Рассмотрен случай структурных ограничений, задаваемых неравенствами. Приведены результаты сравнения решений задачи нахождения стратегии, максимизирующей эффективность функционирования системы, для разных случаев ограничений.

В подразделе 6.3 приведено описание предложенного матричного метода линеаризации систем уравнений в частных производных на постоянных решениях. Приведен пример такой линеаризации, применённой к трехмерным уравнениям газовой динамики с квазигидродинамической регуляризацией, которая состоит из пяти уравнений с пятью неизвестными.

Практически все приведенные результаты были доложены и обсуждены на семинарах, представлены (в этом году онлайн) на конференциях различного уровня. Большинство результатов уже опубликованы, направлены на публикацию или подготовлены к публикации.

Проект 2020г. «Принятия решений и анализ данных в социально-экономических и политических системах» выполнялся в следующих научных направлениях:

- исследование моделей выбора решений и анализа предпочтений;
- исследование моделей сетевого анализа данных;
- исследование по анализу текстов и кластеризации;
- исследования моделей по оптимальному размещению объектов и привлечению частных инвестиций;
- анализ данных и принятие решений в задачах социально-экономического развития регионов;
- исследование моделей регулирования, распространения и специализации.

## **1 Исследование моделей выбора решений и анализа предпочтений**

В этом разделе отчета приведены результаты исследований моделей выбора решений и анализа предпочтений. Работа проводилась в следующих направлениях:

- разработка моделей средних значений выборок в рамках теории важности критериев;
- разработка и анализ моделей расчета степени манипулируемости правил агрегирования предпочтений;
- нахождение вероятностей однопиковых предпочтений;
- исследование турнирных способов выбора оптимальных вариантов;
- анализ предпочтений студентов относительно форматов онлайн-обучения.

### **1.1 Разработка моделей средних значений выборок в рамках теории важности критериев**

#### **1.1.1 Введение**

В управлении, экономике, социологии, технике и других областях науки и практики широко применяются средние величины. В общей теории статистики это понятие определяется таким образом: «Средней величиной называется статистический показатель, который дает обобщенную характеристику варьирующего признака однородных единиц совокупности. <...> Сущность средней заключается в том, что в ней взаимопогашаются случайные отклонения значений признака и учитываются изменения, вызванные основным фактором» [2]. И еще: «Средняя величина – это обобщающая характеристика единиц совокупности по какому-либо варьирующему признаку <...>. Средние величины позволяют сравнивать уровни одного и того же признака в различных совокупностях и находить причины этих расхождений» [3].

Однако следует иметь в виду важное обстоятельство: «... не существует возможности нахождения некой универсальной формулы, исчерпывающей понятие средней величины и обладающей конструктивными достоинствами». (Из предисловия в книге [4].) Поэтому актуальной остается проблема поиска

подходов к общей формулировке понятия средней величины и ее конкретизации для различных ситуаций (задач). В работе предлагаются новые идеи к решению указанной проблемы, основанные на методах многокритериальной оптимизации.

### 1.1.2 Сведения о средних величинах

Вначале для удобства читателя приведем необходимые сведения о средних величинах. Пусть имеется совокупность  $X$ , состоящая из  $n \geq 2$  действительных чисел, называемых далее данными, или точками и являющихся результатами измерения интенсивности некоторого выделенного признака (формула (1.1)) :

$$X = \{x_1, x_2, \dots, x_n\}. \quad (1.1)$$

Эти данные являются однородными в том смысле, что измерения производились по одной и той же шкале [5], характеризуемой множеством допустимых преобразований  $\Phi$  – числовых функций  $\varphi$ , так что вместо множества (1.2) можно рассматривать множество

$$\varphi(X) = \{\varphi(x_1), \varphi(x_2), \dots, \varphi(x_n)\} \quad (1.2)$$

для любой функции  $\varphi \in \Phi$ . Для основных типов шкал [5] областью определения функций  $\varphi$  является вся числовая прямая  $\mathbb{R} = (-\infty, +\infty)$  или же положительный луч  $\mathbb{R}_+ = [0, +\infty)$ .

Упорядоченные соответственно по неубыванию и невозрастанию множества

$$X_{\uparrow} = \langle x_{(1)}, x_{(2)}, \dots, x_{(n)} \rangle; \quad X_{\downarrow} = \langle x_{[1]}, x_{[2]}, \dots, x_{[n]} \rangle, \quad (1.3)$$

где  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$  и  $x_{[1]} \geq x_{[2]} \geq \dots \geq x_{[n]}$ , получаются из совокупности чисел (1.1) при помощи соответствующих перестановок. Отметим, что множество (1.1) в статистике называется выборкой, а первая совокупность из множества (1.3) – вариационным рядом.

Функция  $n$  переменных  $g(X) = f(x_1, x_2, \dots, x_n)$  называется *средней величиной, средней* или *средним (по Коши)*, если (выражение (1.4))

$$\min\{x_1, x_2, \dots, x_n\} \leq g(x_1, x_2, \dots, x_n) \leq \max\{x_1, x_2, \dots, x_n\}. \quad (1.4)$$

Иногда дополнительно требуется, чтобы оба неравенства в выражении (1.4) были строгими [4]. Поскольку все данные в признаке (1.1) считаются «одинаково весомыми», то функция  $g$  должна быть симметрической, так что, в частности,  $g(X) = g(X\uparrow) = g(X\downarrow)$ . Функция  $f$  может задаваться аналитически, алгоритмически или иными способами. Она может быть как однозначной, так и многозначной, и средняя называется соответственно однозначной или многозначной. Адекватный вид средней, т. е. тот конкретный вид функции  $g$ , которую можно корректно применять, зависит от типа шкалы, по которой производились измерения для получения признака (1.1) [6].

Вначале рассмотрим случай качественных данных, когда шкала порядковая, а множество  $\Phi = \Phi_{\Pi}$  включает все возрастающие непрерывные на  $\mathbb{R}$  функции:

$$\Phi_{\Pi} = \{\varphi: x > y \Rightarrow \varphi(x) > \varphi(y); \lim_{x \rightarrow a} \varphi(x) = \varphi(a)\}.$$

Преобразования данных (1.1) с учётом  $\varphi \in \Phi_{\Pi}$  сохраняет их упорядочение:  $x_i > x_j \Rightarrow \varphi(x_i) > \varphi(x_j)$ . Поэтому адекватной средней для порядковой шкалы является, например, медиана  $\mu_X$ , которая определяется с помощью вариационного ряда так: если число  $n$  нечетное, то медиана равна центральному члену ряда:  $\mu_X = x_{(\frac{n+1}{2})}$ ; если же число  $n$  четное, то медианой является любое число, заключенное между значениями двух центральных членов  $x_{(\frac{n}{2})}$  и  $x_{(\frac{n}{2}+1)}$ , так что медиана оказывается многозначной средней.



Будем теперь полагать, что данные (1.1) являются количественными, т. е. измерения производились в шкале интервалов ( $\Phi = \Phi_{\text{и}}$ ) или же шкале отношений ( $\Phi = \Phi_{\text{о}}$ ):

$$\Phi_{\text{и}} = \{ \varphi: \varphi(x) = kx + l, k > 0 \}, \Phi_{\text{о}} = \{ \varphi: \varphi(x) = kx, k > 0 \}.$$

Множество  $\Phi_{\text{и}}$  состоит из всех возрастающих линейных функций, а множество  $\Phi_{\text{о}}$  – из всех возрастающих линейных однородных функций. Для таких данных медиана, разумеется, тоже является адекватной средней. Более того, здесь в случае четного числа  $n$  для устранения многозначности часто за медиану принимают среднюю арифметическую центральных членов  $x_{(\frac{n}{2})}$  и  $x_{(\frac{n}{2}+1)}$  или же иное число, лежащее между ними [4].

В случаях, когда все числа в данных (1.1) неотрицательны, а шкала измерений является шкалой отношений, на практике широко применяется средняя степенная (выражение (1.5))

$$g^s(X) = \left( \frac{1}{n} \sum_{i=1}^n (x_i)^s \right)^{1/s}, s \neq 0. \quad (1.5)$$

Отметим следующие свойства этой средней [4]:

- $g^s(X)$  непрерывна по совокупности переменных  $x_i$  на  $\text{Re}_+^n$  и непрерывна по  $s$  на  $\text{Re}$ , кроме точки  $s = 0$ .
- $g^s(X)$  возрастает по каждой переменной  $x_i$  и возрастает по  $s$ , если не все  $x_i$  равны.

Справедливы формулы (1.6):

$$\lim_{s \rightarrow -\infty} g^s(X) = \min_{i \in N} x_i, \lim_{s \rightarrow 0} g^s(X) = \left( \prod_{i=1}^n x_i \right)^{1/n}, \lim_{s \rightarrow +\infty} g^s(X) = \max_{i \in N} x_i, \quad (1.6)$$

где  $N = \{1, 2, \dots, n\}$ . Если доопределить по непрерывности функцию  $g^s(X)$  для  $s = 0$ , положив ее равной, с учетом второго равенства из (1.6), средней геометрической  $g^0(X) = (\prod_{i=1}^n x_i)^{1/n}$ , то функция  $g^s(X)$  при изменении  $s$  от  $-\infty$  до  $+\infty$  будет пробегать все значения из интервала  $(\min_{i \in N} x_i, \max_{i \in N} x_i)$ .

При  $s = 1$  средняя степенная  $g^s(X)$  оказывается равной средней арифметической  $g^1(X) = \frac{1}{n} \sum_{i=1}^n x_i$ , которая определена и в том более общем случае, когда числа  $x_i$  в данных (1.1) могут иметь любой знак. Среднюю арифметическую можно корректно применять, когда данные (1.1) получены измерениями по шкале интервалов.

### 1.1.3 Сведения из теории многокритериальной оптимизации

Теперь приведем необходимые сведения из теории многокритериальной оптимизации [7]. Далее изложение опирается на следующую математическую модель (формула (1.7)) ситуации принятия решения по  $n \geq 2$  критериям:

$$\langle V, f, R \rangle. \quad (1.7)$$

где  $V$  – множество *вариантов* (действий), или стратегий, планов, альтернатив (оно содержит не менее двух вариантов);  $f = (f_1, f_2, \dots, f_n)$  – *векторный критерий*, состоящий из  $n$  частных критериев  $f_i: V \rightarrow Z_i$ , т. е. *критерий*  $f_i$  – это функция с областью определения  $V$  и числовой областью значений  $Z_i \subseteq \mathbb{R}$ . Каждый вариант  $v$  из множества  $V$  характеризуется своей *векторной оценкой*  $y(v) = f(v) = (f_1(v), f_2(v), \dots, f_n(v))$ . Поэтому сравнение вариантов по предпочтительности сводится к сопоставлению их векторных оценок. Множество всех векторных оценок есть  $Z = Z_1 \times Z_2 \times \dots \times Z_n$ . Через  $Y$  обозначается множество достижимых векторных оценок, т.е. множество значений векторного критерия:  $Y = f(V) = \{y \in Z \mid \exists v \in V: y = f(v)\}$ .

Предпочтения моделируются на множестве  $Z$  при помощи *отношения нестрогого предпочтения*  $R$ , так что  $yRz$  означает, что векторная оценка  $y$  не

менее предпочтительна, чем оценка  $z$ . Отношение  $R$  является *квазипорядком*, т. е. оно рефлексивно и транзитивно: верно  $yRu$  и  $(yRz, zRu)$  влечет  $yRu$  при любых  $y, z, u \in Z$ . Если найдутся две векторные оценки  $y$  и  $z$ , несравнимые по  $R$ , т. е. для которых неверно ни  $yRz$ , ни  $zRu$ , то квазипорядок называется *частичным*, или *несвязным*, в противном случае – *полным*, или *связным*.

Отношение  $R$  порождает *отношение безразличия*  $I$  и (строгого) *предпочтения*  $P$  следующим образом:  $yIz \Leftrightarrow yRz \wedge zRu$ ;  $yPz \Leftrightarrow yRz \wedge zRu$  (здесь  $zRu$  означает, что  $zRu$  неверно). Отношение  $R$  определяется составом информации о предпочтениях ЛПР. Запись  $yIz$  означает, что векторные оценки  $y$  и  $z$  одинаковы по предпочтительности, или безразличны, а запись  $yPz$  означает, что векторная оценка  $y$  предпочтительнее, чем оценка  $z$ . Отношение  $R$ , заданное на множестве  $Z$  подходящим решающим правилом на основе имеющейся информации  $\Gamma$  о предпочтениях ЛПР, обозначается  $R^\Gamma$ . Отношения  $R^\Gamma$ ,  $I^\Gamma$  и  $P^\Gamma$  порождают имеющие аналогичный смысл отношения  $R_\Gamma$ ,  $I_\Gamma$  и  $P_\Gamma$  на множестве  $V$  (выражение (1.8)):

$$vR_\Gamma w \Leftrightarrow f(v)R^\Gamma f(w), vI_\Gamma w \Leftrightarrow f(v)I^\Gamma f(w), vP_\Gamma v \Leftrightarrow f(v)P^\Gamma f(w). \quad (1.8)$$

(Достижимая) векторная оценка  $y^* \in Y$  называется *недоминируемой* (по  $P^\Gamma$ ), если не существует (достижимой) векторной оценки  $y \in Y$  такой, что верно  $yP^\Gamma y^*$ ; в противном случае она – *доминируемая* (по отношению  $P^\Gamma$  векторной оценкой  $y$ ). Недоминируемым векторным оценкам соответствуют недоминируемые варианты: вариант  $v^*$  недоминируем (по  $P_\Gamma$ ), если не существует варианта  $v$  такого, что верно  $vP_\Gamma v^*$ . Пусть  $Y^\Gamma$  и  $V^\Gamma$  – множества недоминируемых векторных оценок и вариантов соответственно. Любые две недоминируемые векторные оценки (недоминируемые варианты) либо безразличны, либо несравнимы. Множество  $Y^\Gamma$  (множество  $V^\Gamma$ ) называется *внешне устойчивым*, если для всякой доминируемой векторной оценки  $y$  (доминируемого варианта  $v$ ) найдется недоминируемая векторная оценка  $y^*$  (недоминируемый вариант  $v^*$ ) такая (такой), что верно  $y^*P^\Gamma y$  (соответственно  $v^*P_\Gamma v$ ). Оптимальный

(наилучший) вариант надлежит выбрать среди множества недоминируемых вариантов, если оно внешне устойчиво.

Множества  $R_+^\Gamma(y) = \{z \in A \mid zR^\Gamma y\}$  и  $R_-^\Gamma(y) = \{z \in A \mid yR^\Gamma z\}$  называются соответственно верхним и нижним срезами множества  $A \subseteq Z$  отношением  $R^\Gamma$  через точку  $y \in A$ . Если верхний срез через каждую точку  $y \in Y$  является множеством замкнутым и ограниченным, то множество недоминируемых векторных оценок  $Y^\Gamma$  не пусто и, более того, внешне устойчиво. Полезно иметь в виду, что если множество стратегий  $V \subset \mathbb{R}^m$  является замкнутым и ограниченным, а все критерии  $f_i$  – непрерывные на  $V$  функции, то и множество  $V^\Gamma$  замкнуто и ограничено.

Будем далее полагать, что предпочтения ЛПП с увеличением значений критериев  $f_i$  убывают, или, иными словами, что критерии желательно минимизировать<sup>1</sup>. При отсутствии иной информации о предпочтениях на множестве  $Z$  предпочтения описывает отношение Парето  $P^\emptyset$ , определяемое так (выражение (1.9)):

$$yP^\emptyset z \Leftrightarrow (y_i \leq z_i, i = 1, 2, \dots, n; y \neq z). \quad (1.9)$$

Отношение  $R_\emptyset$  определяется  $n$  нестрогими неравенствами из (1.9). Отношение  $P^\emptyset$  порождает на  $V$  отношение Парето  $P_\emptyset$ :  $vP_\emptyset w \Leftrightarrow f(v)P^\emptyset f(w)$ .

Недоминируемые по  $P^\emptyset$  векторные оценки (и недоминируемые по соответствующему отношению  $P_\emptyset$  варианты) называются оптимальными по Парето и составляют множество  $Y^\emptyset$  (соответственно  $V_\emptyset$ ). Определения отношения  $P_\emptyset$  и множества  $V_\emptyset$  корректно, даже если каждый критерий имеет свою

---

<sup>1</sup> Это следует помнить при обращении к указанной в статье литературе, так как в ней предполагается, что критерии желательно максимизировать.

всего лишь порядковую шкалу: после применения к критериям  $f_i$  соответствующих допустимых преобразований  $\varphi_i$  (возрастающих на  $\mathbb{R}$  непрерывных функций) указанные отношение и множество не изменяются.

Теперь предположим, что все критерии имеют общую порядковую шкалу, так что  $Z_1 = Z_2 = \dots = Z_n = Z_0$ , и  $Z = Z_0^n$ . Рассмотрим случай, когда все критерии равноважны [8]. Пусть  $\Pi$  – множество перестановок  $\pi = \langle \pi(1), \pi(2), \dots, \pi(n) \rangle$  множества  $\{1, 2, \dots, n\}$ . Критерии  $f_1, f_2, \dots, f_n$  называются *равноважными* (имеющими одинаковую важность), если одинаковы по предпочтительности всякая векторная оценки  $y$  и полученная из нее при помощи любой перестановки  $\pi \in \Pi$  векторная оценка  $\pi(y) = (y_{\pi(1)}, y_{\pi(2)}, \dots, y_{\pi(n)})$ . Информацию о том, что все критерии равноважны, будем обозначать через  $E$ . Отношение нестрогого предпочтения  $R^E$  (квазипорядок), порождаемое на  $Z$  такой информацией, задается следующим образом (выражение (1.10)):

$y R^E z \Leftrightarrow$  найдутся такие перестановки  $\pi, \rho \in \Pi$ , что

$$y_{\pi(1)} \leq z_{\rho(1)}, y_{\pi(2)} \leq z_{\rho(2)}, \dots, y_{\pi(n)} \leq z_{\rho(n)}; \quad (1.10)$$

при этом  $y I^E z$ , когда все  $n$  нестрогих неравенств выполняются как равенства, и  $y P^E z$ , когда хотя бы одно из неравенств является строгим. Эти отношения порождают, согласно выражению (1.8), соответствующие отношения  $R_E, I_E$  и  $P_E$  на множестве вариантов  $V$ .

Следующие равносильные утверждения задают простые правила проверки справедливости выполнения соответствующих соотношений (выражение (1.3)):

$$y R^E z \Leftrightarrow y_{(1)} \leq z_{(1)}, y_{(2)} \leq z_{(2)}, \dots, y_{(n)} \leq z_{(n)};$$

$$y R^E z \Leftrightarrow y_{[1]} \leq z_{[1]}, y_{[2]} \leq z_{[2]}, \dots, y_{[n]} \leq z_{[n]}. \quad (1.11)$$

И здесь также справедливы замечания, аналогичные приведенным сразу после выражения (1.10).

Отношения  $P^E$  и  $P_E$  порождают соответствующие множества  $Y^E$  и  $V_E$  недоминируемых соответственно по  $P^E$  и  $P_E$  векторных оценок и вариантов. Определения отношений  $R_E, P_E, I_E$  и множества  $V_E$  корректны, даже если общая шкала критериев всего лишь порядковая.

Заданная на  $Z$  симметрическая функция  $\psi(y)$  называется возрастающей (неубывающей, убывающей, невозрастающей) по  $P^E$ , если из  $y P^E z$  следует  $\psi(y) > \psi(z)$  (соответственно  $\psi(y) \geq \psi(z)$ ,  $\psi(y) < \psi(z)$ ,  $\psi(y) \leq \psi(z)$ ). Примером такой убывающей по  $P^E$  функции является  $\psi_\Sigma(y) = \sum_{i=1}^n y_i$ , а невозрастающей – функция  $\psi^\bullet(y|a) = \min_{\pi \in \Pi} \max_{i \in N} \{y_{\pi(i)} - a_i\}$ , где  $a$  – произвольный вектор из  $Z$ .

Пусть  $y^*$  – точка минимума функции  $\psi(y)$ , неубывающей по  $P^E$  на  $Z$ . Для того чтобы точка  $y^*$  была недоминируемой по  $P^E$ , достаточно, чтобы функция  $\psi$  была убывающей по  $P^E$  или чтобы эта точка была единственной (с точностью до эквивалентности  $I^E$ ) точкой минимума указанной функции. Если  $y^*$  – недоминируемая по  $P^E$  точка, то она является единственной (с точностью до эквивалентности  $I^E$ ) точкой минимума функции  $\psi^\bullet(y|y^*)$  на  $Y$ , причем этот минимум равен нулю. Следовательно, все недоминируемые по  $P^E$  векторные оценки будут найдены в результате решения параметрической задачи минимизации функции  $\psi^\bullet(y|a)$  для  $a \in Y$  (но при этом могут быть выделены и доминируемые).

Выделить оптимальные варианты среди недоминируемых по  $P_E$  позволяет дополнительная информация о предпочтениях, касающаяся, в частности, возможности взаимной компенсации значений критериев. Если увеличение больших значений одних критериев не компенсируется уменьшением меньших значений других, то предпочтения на множестве  $Z$  описываются лексикографическим отношением  $R^{EL}$  (выражение 1.12)):

$$y R^{EL} z \Leftrightarrow (y_{[1]} < z_{[1]}) \vee (y_{[1]} = z_{[1]}, y_{[2]} < z_{[2]}) \vee \dots \vee (y_{[i]} = z_{[i]}, i = 1, 2, \dots, n), \quad (1.12)$$

причем  $yP^{EL}z$  верно, если последние  $n$  равенств в (1.12) не выполнены. Отношение  $R^{EL}$  – полный квазипорядок. Векторная оценка  $y^* \in Y$  оптимальна по  $R^{EL}$ , если  $y^*R^E y$  для любой  $y \in Y$ . Отношение  $R^{EL}$  порождает на  $V$  отношение  $R_{EL}$ . Все оптимальные векторные оценки (варианты) эквивалентны по  $I^E$  (соответственно по  $I_E$ ) и недоминируемы по  $P^E$  (по  $P_E$ ).

Пусть, наконец, известно, что общая шкала равноважных критериев является шкалой первой порядковой метрики [9]. Это означает, что если в произвольной векторной оценке  $y$ , в которой  $y_i > y_j$ , заменить  $y_i$  на  $y_i - \delta$ , а  $y_j$  на  $y_j + \delta$ , где  $\delta$  – положительное число такое, что  $y_i - \delta \geq y_j + \delta$ , то полученная таким образом векторная оценка  $z$  будет предпочтительнее, чем исходная  $y$ . Отношение нестрогого предпочтения  $R^{E\Delta}$ , порождаемое такой информацией на  $Z$ , задается так [10], [11]:

$$yR^{E\Delta}z \Leftrightarrow (y_{[1]} \leq z_{[1]}, y_{[1]} + y_{[2]} \leq z_{[1]} + z_{[2]}, \dots, y_{[1]} + y_{[2]} + \dots + y_{[n]} \leq z_{[1]} + z_{[2]} + \dots + z_{[n]}).$$
(1.13)

И здесь справедливы замечания, аналогичные приведенным сразу после выражения (1.10).

#### 1.1.4 Определение средних величин как недоминируемых точек

Рассмотрим множество данных  $X$  (выражение (1.1)). Пусть  $x$  – произвольное фиксированное число – точка на числовой прямой  $\text{Re}$ . Удаленность ее от отдельной точки  $x_i$  из  $X$  можно оценить расстоянием  $y_i = |x - x_i|$ . Тогда удаленность  $x$  от совокупности всех точек из  $X$  характеризуется вектором  $y = (y_1, y_2, \dots, y_n)$ , составленным из таких расстояний. Его можно считать значением векторного критерия  $f(x) = (f_1(x), f_2(x), \dots, f_n(x))$ , где  $f_i(x) = |x - x_i|$ . Областью значений  $Z$  этого векторного критерия является положительный квадрант  $\text{Re}_+^n = [0, +\infty)^n$  – множество векторов с неотрицательными компонентами,

а множество значений  $Y = f(X) \subset Z$  может иметь довольно сложную конфигурацию уже при  $n = 3$ . Для случаев  $n = 2$  и  $n = 3$  оно изображено на рисунках 1.1 и 1.2.

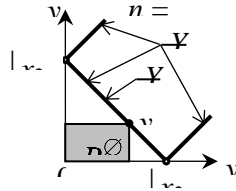


Рисунок 1.1 – График множества векторных оценок  $Y$

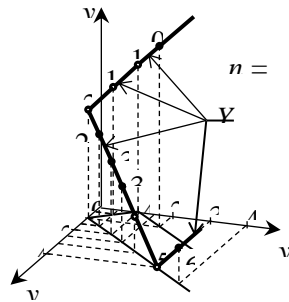


Рисунок 1.2 – График множества векторных оценок  $Y$

Пусть на множестве  $Z$  задано отношение нестрого предпочтения – частичный квазипорядок  $R^\Gamma$ , где  $\Gamma$  – информация о предпочтениях, касающаяся удаленности: если верно  $xP_\Gamma x'$ , то точка  $x$  ближе к совокупности точек  $X$ , чем  $x'$ . Следовательно, на роль наиболее близких к  $X$  и представляющих все множество  $X$  могут претендовать лишь те точки, которые недоминируемы по  $R_\Gamma$ . Если множество таких точек  $G^\Gamma(X)$  внешне устойчиво, то все они будут именоваться ПН-средними, а более конкретно (для рассматриваемой информации  $\Gamma$ ) и кратко – средними по  $R_\Gamma$ .

**Теорема 1.1.** *Средними по  $P_\emptyset$  являются все точки отрезка от  $x_{(1)} = \min_{i \in N} x_i$  до  $x_{(n)} = \max_{i \in N} x_i$ , т. е.  $G^\emptyset(X) = \bar{X} = [x_{(1)}, x_{(n)}]$ . Это множество внешне устойчиво.*



Таким образом, понятие средних по  $P_\emptyset$  оказывается эквивалентным понятию средних по Коши.

На рисунке 1.1 множество  $Y^\emptyset$  векторных оценок средних по  $P_\emptyset$  – это один отрезок. На рисунке 1.2 множество  $Y^\emptyset$  состоит из двух звеньев ломаной – от  $(0, 1, 4)$  до  $(1, 0, 3)$  и от  $(1, 0, 3)$  до  $(4, 3, 0)$ .

Множество  $G^\emptyset(X)$  имеет очень простую структуру и задается явно. Однако для других отношений  $P_\Gamma$  сложность конфигурации множества  $Y^\Gamma$  приводит к сложности конфигурации и структуры множества ПН-средних  $G^\Gamma(X)$ , что существенно затрудняет его построение (см. далее).

Отметим, что если  $\varphi$  – возрастающая на  $\text{Re}_+$  функция, то замена исходных критериев  $f_i(x) = |x - x_i|$  на критерии  $\varphi(f_i(x))$  не изменяет множество  $G^\emptyset(X)$ . Например, можно перейти к «гладким» критериям  $f_i(x) = (x - x_i)^2$ .

#### 1.1.5 Средние величины при равноважных критериях

Далее будем полагать, что все критерии имеют равную важность (информация  $E$ ). В этом случае удаленность точки  $x$  от множества  $X$  оценивается отношением  $R_E$  на  $\text{Re}^n$ , которое определяется согласно соотношениям (1.11) каждым из двух равносильных правил:

$$xR_{Ex'} \Leftrightarrow f_{(1)}(x) \leq f_{(1)}(x'), f_{(2)}(x) \leq f_{(2)}(x'), \dots, f_{(n)}(x) \leq f_{(n)}(x');$$

$$xR_{Ex'} \Leftrightarrow f_{[1]}(x) \leq f_{[1]}(x'), f_{[2]}(x) \leq f_{[2]}(x'), \dots, f_{[n]}(x) \leq f_{[n]}(x'), \quad (1.14)$$

где  $f_i(x) = |x - x_i|$ . ПН-средними (по  $R_E$ ), составляющими  $G^E(X)$ , здесь являются недоминируемые по  $P_E$  точки числовой прямой.

**Теорема 1.2.** *Множество  $G^E(X) \subseteq G^\emptyset(X) = \bar{X}$  и внешне устойчиво.*

Отметим, что если  $\varphi$  – возрастающая на  $\text{Re}_+$  функция, то замена всех исходных критериев  $f_i(x)$  на критерии  $\varphi(f_i(x))$  не изменяет отношение  $R_E$  и множество  $G^E(X)$ .

**Пример 1.1.** Нетрудно понять, что при  $n = 2$  любые две различные точки из  $\bar{X}$  не сравнимы по отношению  $P_E$  и поэтому имеет место равенство  $G^E(X)$

$= \bar{X}$ . Заметим, что верхние срезы  $\text{Re}_+^n$  отношением  $P^E$  не выпуклы (рисунок 1.3).

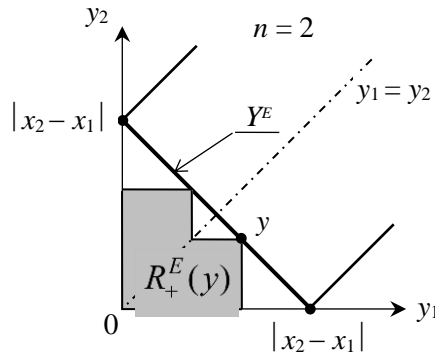


Рисунок 1.3 – График множества векторных оценок  $Y^E$

Пример 1.2. Пусть  $n = 3$  и  $X = \{1, 2, 5\}$ . Здесь  $G^E(X) = [1, 5; 3]$ . На рис. 1.2 множество  $Y^E$  состоит из двух звеньев ломаной – от  $(\frac{1}{2}, \frac{1}{2}, 3\frac{1}{2})$  до  $(1, 0, 3)$  и от  $(1, 0, 3)$  до  $(2, 1, 2)$ .

В этих двух примерах структура множества  $G^E(X)$  проста – это один отрезок. Однако при больших значениях  $n$  это множество может представлять собой объединение нескольких промежутков, границы которых могут им не принадлежать! Это – следствие невыпуклости и замкнутости верхних срезов  $\text{Re}_+^n$  отношением  $P^E$ .

Пример 1.3. При  $n = 6$  имеем:

$$G^E(\{10, 11, 15, 61, 107, 110\}) = [10, 5; 83) \cup (83, 5; 84, 5) \cup (106, 5; 109);$$

$$G^E(\{10, 11, 40, 55, 70, 110\}) = [10, 5; 18) \cup (18; 67, 5) \cup (68; 75);$$

$$G^E(\{10, 57, 61, 64, 109, 110\}) = (56, 5; 57, 5) \cup (58, 5; 88, 5) \cup (108; 109, 5].$$

Примеры 1.2 и 1.3 иллюстрируют и следующую теорему.

**Теорема 1.3.** Если  $x_{(i)}$  и  $x_{(i+1)}$  – два неравных смежных члена вариационного ряда с минимальным расстоянием  $x_{(i)} - x_{(i+1)}$  между ними и такой интервал единствен, то их середина – точка  $x^c = \frac{1}{2}(x_{(i)} + x_{(i+1)})$  входит в состав  $G^E(X)$ ; при этом если  $x_{(i)}$  есть  $x_{(1)}$  (соответственно  $x_{(i+1)}$  есть  $x_{(n)}$ ), то  $x^c$  – это левая (соответственная правая) точка множества  $G^E(X)$ .

Не при всяком значении параметра  $s$  средняя степенная  $g^s(X)$  (см. формулу (1.5)) является и средней по  $P_E$ , если  $x_{(1)} \neq x_{(n)}$  и  $(x_{(1)}, x_{(n)}) \not\subset G^E(X)$ , ибо продолженная по непрерывности функция  $g^s(X)$  при возрастании  $s$  на  $\mathbb{R}_E$  пробегает все значения из интервала  $(x_{(1)}, x_{(n)})$ . Однако справедлива

**Теорема 1.4.** *Средняя арифметическая является и средней по  $P_E$ , т.е.  $g^1(X) \in G^E(X)$ .*

**Пример 1.4.** Согласно примеру 3 для  $X = \{10, 57, 61, 64, 109, 110\}$  имеем:  $G^E(X) = (56,5; 57,5) \cup (58,5; 88,5) \cup (108; 109,5]$ ; здесь средняя геометрическая  $g^0(X) = 54,66 \notin G^E(X)$  и средняя гармоническая  $g^{-1}(X) = 35,75 \notin G^E(X)$ , но, конечно,  $g^1(X) = 68,5 \in G^E(X)$ . В примере 2 для  $X = \{1, 2, 5\}$  имеем  $G^E(X) = [1,5; 3]$ ; здесь средняя квадратическая  $g^2(X) = 3,162 \notin G^E(X)$ , но  $g^1(X) = 2,67 \in G^E(X)$ .

**Теорема 1.5.** *Медиана является и средней по  $P_E$ , т.е. при нечетном  $n$ , когда медиана однозначна, верно  $\mu(X) = x_{(\frac{n+1}{2})} \in G^E(X)$ , а при нечетном  $n$ , когда медиана многозначна, верно  $\mu(X) = [x_{(\frac{n}{2})}, x_{(\frac{n}{2}+1)}] \subseteq G^E(X)$ .*

Это утверждение иллюстрируют примеры 1.2 и 1.3.

#### 1.1.6 Средние величины при равноважных критериях и отсутствии компенсации удаленностей

Пусть увеличения удаленности точки  $x$  от одних точек  $x_i$  не компенсируются уменьшением ее удаленности от других. В этом случае для определения ПН-средних можно воспользоваться отношением  $R_{EL}$ , задаваемым на  $X$  правилом согласно отношению (1.12):

$$xR_{EL}x' \Leftrightarrow [f_{[1]}(x) < f_{[1]}(x')] \vee [f_{[1]}(x) = f_{[1]}(x'), f_{[2]}(x) < f_{[2]}(x')] \vee \dots \\ \dots \vee [f_{[i]}(x) = f_{[i]}(x'), i = 1, 2, \dots, n]. \quad (1.15)$$

Поскольку расстояние от середины отрезка  $\bar{X}$  — точки  $x^c = \frac{1}{2}(x_{[1]} + x_{[n]})$  — до каждой из двух крайних точек множества  $X$  равно  $d = \frac{1}{2}(x_{[1]} - x_{[n]})$ , а для любой другой точки  $x$  расстояние до одной из крайних точек больше, чем  $d$ , то

точка  $x^c$  является оптимальной по  $R_E$ , т. е. верно  $x^c R_{EL} x$ . Следовательно,  $x^c$  является однозначной средней по  $R_{EL}$  и  $G^{EL}(X) = \{x^c\}$ . Так как  $P_{EL} \supset P_E$ , то  $x^c \in G^E(X)$ , т. е. в число средних по  $R_E$  всегда входит середина отрезка  $\bar{X}$ .

### 1.1.7 Средние величины при равноважных критериях со шкалой первой порядковой метрики

Будем рассматривать случай, когда векторная оценка  $z$ , полученная из произвольной векторной оценки  $y$ , в которой  $y_i > y_j$ , заменой  $y_i$  на  $y_i - \delta$  и  $y_j$  на  $y_j + \delta$ , где  $\delta$  – положительное число такое, что  $y_i - \delta \geq y_j + \delta$ , будет предпочтительнее, чем исходная  $y$  в том смысле, что она, как более «сконцентрированная», считается более подходящей для роли средней (такая информация будет обозначаться буквой  $\Delta$ ). Это означает, что общая шкала равноважных критериев является шкалой первой порядковой метрики. Отношение нестрогого предпочтения  $R_{E\Delta}$ , порождаемое совокупной информацией  $E$  и  $\Delta$  на  $\text{Re}^n$ , задается, согласно отношению (1.13), таким решающим правилом:

$$\begin{aligned} x R_{E\Delta} x' \Leftrightarrow & [f_{[1]}(x) \leq f_{[1]}(x'), f_{[1]}(x) + f_{[2]}(x) \leq f_{[1]}(x') + f_{[2]}(x'), \dots \\ & \dots f_{[1]}(x) + f_{[2]}(x) + \dots + f_{[n]}(x) \leq f_{[1]}(x') + f_{[2]}(x') + \dots + f_{[n]}(x')]. \end{aligned} \quad (1.16)$$

ПН-средними здесь являются точки, недоминируемые по  $P_{E\Delta}$ . Так как  $P_{E\Delta} \supset P_E$ , то  $G^E(X) \supseteq G^{E\Delta}(X)$ . Отметим, что верхние срезы  $\text{Re}_+^n$  отношением  $P^{E\Delta}$  – выпуклые множества (рисунок 1.4). Так как  $P_{EL} \supset P_{E\Delta}$ , то  $G^E(X) \supseteq G^{EL}(X) = \{x^c\}$ .

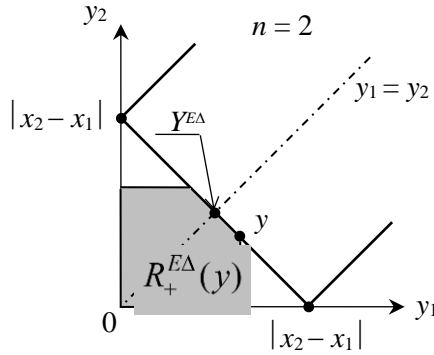


Рисунок 1.4 – График множества векторных оценок  $Y^{E\Delta}$

Теорема 1.6. Средняя арифметическая является и средней по  $P_{E\Delta}$ , т. е.  $g^1(X) \in G^{E\Delta}(X)$ .

Теорема 1.7. При нечетном  $n$ , когда медиана однозначна, она является и средней по  $P_{E\Delta}$ , т. е. верно  $\mu(X) \in G^{E\Delta}(X)$ ; при четном  $n$ , когда медиана многозначна, верно лишь  $\mu(X) \cap G^{E\Delta}(X) \neq \emptyset$ .

Пример 1.5. При  $n = 5$  для  $X = \{1, 2, 3, 5, 11\}$  имеем  $G^{E\Delta}(X) = [3; 6]$ ,  $\mu(X) = 3$ ,  $g^1(X) = 4,4$ . При  $n = 4$  для  $X = \{10, 11, 12, 110\}$  имеем  $G^{E\Delta}(X) = [11,5; 60]$ ,  $\mu(X) = [11; 12]$ ,  $g^1(X) = 35,75$ , а для  $X = \{10, 11, 20, 110\}$  имеем  $G^{E\Delta}(X) = [15,5; 60]$ ,  $\mu(X) = [11; 20]$ ,  $g^1(X) = 35,75$ .

Введем в рассмотрение функции

$$\sigma_k(x) = \sum_{i=1}^k f_{[i]}(x), \quad k = 1, 2, \dots, n. \quad (1.17)$$

Теорема 1.8. Функции (1.17) являются выпуклыми кусочно-линейными, их графики – ломаные линии.

С помощью функций (1.17) решающее правило (1.16) можно переписать в таком виде:

$$x R_{E\Delta} x' \Leftrightarrow \sigma_k(x) \leq \sigma_k(x'), \quad k = 1, 2, \dots, n. \quad (1.18)$$

Заметим, что, согласно решающему правилу (1.18), отношение  $R_{E\Delta}$  можно считать отношением Парето (отношением покомпонентного нестрого доминирования) по векторному критерию  $\sigma(x) = (\sigma_1(x), \sigma_2(x), \dots, \sigma_n(x))$ , и множество  $G^{E\Delta}(X)$  считать множеством Парето для векторного критерия  $\sigma(x)$ .

Обозначим области минимальных значений выпуклых кусочно-линейных функций  $\sigma_k(x)$  через  $M_k = \{x : \sigma_k(x) = \min \sigma_k(x)\}$ . Эти области являются отрезками  $M_k = [a_k, b_k]$  (но некоторые отрезки стягиваются в одну точку). Пусть  $\alpha = \min_k b_k$ ,  $\beta = \max_k a_k$ . Так как минимум функции  $\sigma_1(x) = f_{[1]}(x)$  всегда достигается только в середине отрезка  $\bar{X}$  – точке  $x^c = \frac{1}{2}(x_{(1)} + x_{(n)})$ , то  $M_1 = \{x^c\}$ . Поэтому верны соотношения:  $\alpha = \min_k b_k \leq b_1 = x^c = a_1 \leq \max_k a_k = \beta$ , так что  $\alpha \leq \beta$ .

Следующее утверждение показывает, что структура множества  $G^{E\Delta}(X)$  совсем простая.

**Теорема 1.9.** *Множество  $G^{E\Delta}(X)$  внешне устойчиво и является отрезком  $[\alpha, \beta]$ .*

**Пример 1.6.** При  $n = 2$  множество  $G^{E\Delta}(X)$  состоит из одной точки  $x^c$  (рисунки 1.4 – 1.6).

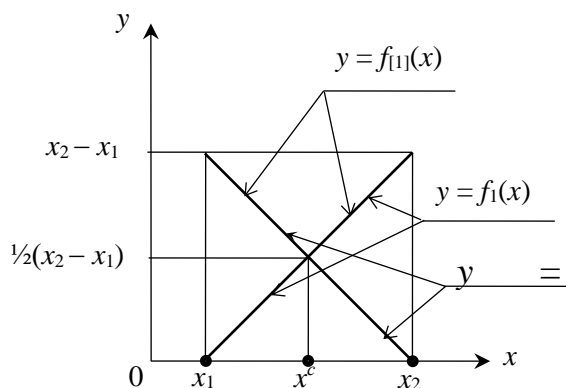


Рисунок 1.5 – Графики функций  $y = f_1(x)$  и  $y = f_2(x)$  на  $\bar{X}$

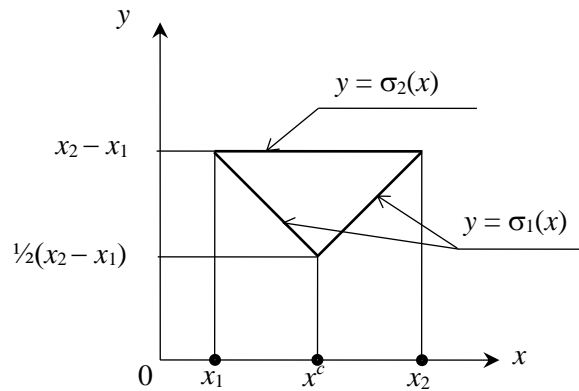


Рисунок 1.6 – Графики функций  $y = \sigma_1(x)$  и  $y = \sigma_2(x)$  на  $\bar{X}$

Пример 1.7. При  $n = 5$  для  $X = \{1, 2, 7, 8, 11\}$  графики функций  $y = \sigma_k(x)$  представлены на рисунок 1.7. Он показывает, что здесь  $M_1 = \{6\}$ ,  $M_2 = [4, 5; 6, 5]$ ,  $M_3 = \{5\}$ ,  $M_4 = [4, 5; 7, 5]$ ,  $M_5 = \{7\}$ ,  $\alpha = 5$ ,  $\beta = 7$ . Поэтому  $G^{E\Delta}(X) = [5; 7]$ .

Пример 1.8. При  $n = 3$  для  $X = \{1, 2, 5\}$  имеем:  $G^{E\Delta}(X) = [2; 3] \subset G^E(X) = [1, 5; 3]$ . При  $n = 5$  для  $X = \{1, 2, 5, 9, 11\}$  имеем:  $G^{E\Delta}(X) = [5; 6] \subset G^E(X) = [1, 5; 7, 5] \cup (8, 5; 9, 5)$ . При  $n = 6$  для  $X = \{10, 11, 15, 61, 107, 110\}$  имеем:  $G^{E\Delta}(X) = [38; 60] \subset G^E(X) = [10, 5; 83] \cup (83, 5; 84, 5) \cup (106, 5; 109)$ .

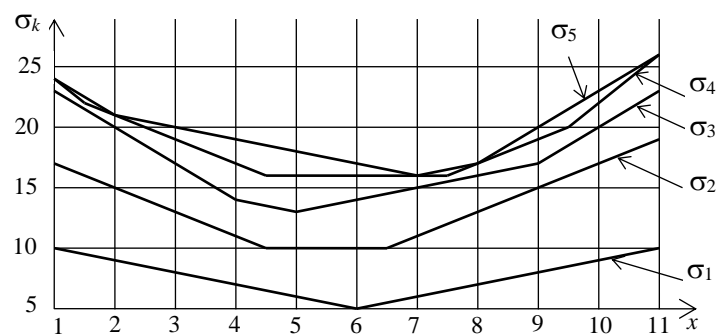


Рисунок 1.7 – Графики функций  $\sigma_k(x)$

### 1.1.8 О построении множеств средних величин

Для построения множества  $G^E(X)$  можно воспользоваться известными в теории многокритериальной оптимизации подходящими методами построения множеств недоминируемых вариантов [7], [8], основанных на применении семейств функций, возрастающих (убывающих) или хотя бы не убывающих (не возрастающих) по  $P_E$ . Например, можно решать параметрическую задачу минимизации на  $X$  функции одной переменной  $\psi^\bullet(f(x)|c) = \min_{\pi \in \Pi} \max_{i \in N} \{f_{\pi(i)}(x) - c_i\}$ , варьируя векторный параметр  $c \in f(\bar{X})$ . Однако уже при «не очень большом» количестве точек  $n$  число  $n!$  членов в этой функции, среди которых ищется максимум, становится слишком большим.

Учитывая одномерность множества  $X$ , можно пойти другим путем: построить равномерную сетку с «небольшим» шагом  $h$ , покрывающую множество  $\bar{X}$ , и искать недоминируемые по  $P_E$  точки среди узлов этой сетки прямым перебором (с помощью известных приемов сокращения объема перебора [12]). Шаг сетки зависит от требуемой точности расчетов и может уменьшаться в процессе их проведения при решении задачи построения  $G^E(X)$ . Именно при помощи такого способа находились множества  $G^E(X)$  в примерах 1.2 и 1.3.

Пример 1.9. Покажем вкратце, как можно решить задачу построения  $G^E(X)$  для  $X = \{1, 2, 5, 9, 11\}$ . Расчеты на компьютере с постепенным уменьшением шага сетки дали следующие результаты построения множества недоминируемых точек:

Шаг $h = 1$ :	$[2; 7] \cup [9; 9]$ .
Шаг $h = 0,1$ :	$[1,5; 7,4] \cup [8,6; 9,4]$ .
Шаг $h = 0,01$ :	$[1,50; 7,49] \cup [8,51; 9,49]$ .
Шаг $0,001$ :	$[1,500; 7,499] \cup [8,501; 9,499]$ .
Шаг $0,0001$ :	$[1,5000; 7,4999] \cup [8,5001; 9,4999]$ .

При переборе с шагом  $0,01$  было выяснено, что точка  $4,5$  доминирует над точками  $7,5$  и  $8,5$ , а точка  $2,5$  доминирует над точкой  $9,5$ . Поэтому, с учетом теоремы 1.2, принято  $G^E(X) = [1,5; 7,5) \cup (8,5; 9,5)$ .



Для нахождения множества  $G^{E\Delta}(X)$  можно воспользоваться аналогом описанного выше метода, основанного на применении равномерной сетки: построить при её помощи графики функций  $y = \sigma_k(x)$ , а затем выделить отрезок  $[\alpha, \beta]$  (см. теорему 1.9).

### 1.1.9 О сравнении множественных средних по величине

Практически важно умение сравнивать средние, полученные при измерениях по одной и той же шкале, по величине. Для однозначных средних такой вопрос не стоит, так как сравнение двух средних сводится просто к сопоставлению двух чисел. Многочисленные средние в статистике обычно «уточняют», заменяя одним числом (пример – медиана для нечетного числа  $n$ ).

Практически важно умение сравнивать средние, полученные при измерениях по одной и той же шкале, по величине. Для однозначных средних такой вопрос не стоит, так как сравнение двух средних сводится просто к сопоставлению двух чисел. Многочисленные средние в статистике обычно «уточняют», заменяя одним числом (пример – медиана для нечетного числа  $n$ ).

Пример 1.10. Пусть  $G^E(X') = [1; 2) \cup (5; 8)$ ,  $G^E(X'') = [1,5; 4,5] \cup (8; 9]$ . Графики функций  $d_x^E(X')$  и  $d_x^E(X'')$  представлены на рисунке 1.8. Поскольку ни один из этих графиков не лежит под другим, то рассматриваемые средние не сравнимы по  $\succeq$ , т.е. это отношение не позволяет утверждать, что одна из средних больше другой или же что они равны.

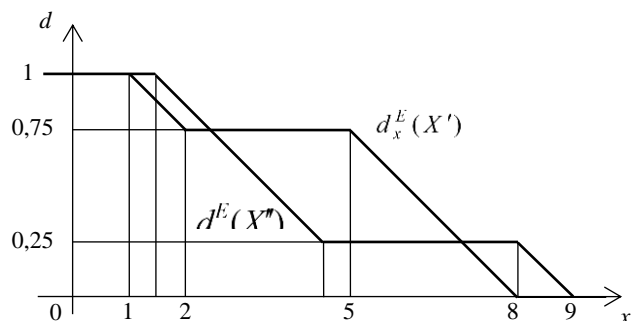


Рисунок 1.8 – Графики функций  $d_x^E(X')$  и  $d_x^E(X'')$

Отношение  $\succeq$  является слабым в том смысле, что редко позволяет сравнить по величине множественные средние. Воспользовавшись идеями стохастического доминирования второго порядка, отношение  $\succeq$  можно расширить, но практически это не очень эффективно.

Можно, разумеется, «сжать» множественные средние до однозначных. Но при этом будет потеряно много информации, т.е. результаты сравнения будут весьма грубыми. Например, среднюю  $G^\Gamma(X)$ , состоящую из  $l$  непересекающихся непустых интервалов с концами  $x^1, x^2; x^3, x^4; \dots; x^{2l-1}, x^{2l}$ , можно представить одной точкой – центром тяжести этих интервалов  $x^\Gamma(X)$ .

Пример 1.11. Для средних из примера 1.8 имеем:

$$x^E(X') = (1,5 \cdot 1 + 6,5 \cdot 3) / 4 = 4,875; x^E(X'') = (3 \cdot 3 + 8,5 \cdot 1) / 4 = 4,375.$$

Поскольку  $4,875 > 4,375$ , то, согласно рассматриваемому подходу, следует принять, что средняя  $G^E(X')$  больше, чем средняя  $G^E(X'')$ .

Полезно иметь в виду, что если  $G^\Gamma(X') \succ G^\Gamma(X'')$ , то  $x^\Gamma(X') > x^\Gamma(X'')$  (см. [13]).

Сравнивать средние  $G^{E\Delta}(X')$  и  $G^{E\Delta}(X'')$  проще, чем средние  $G^E(X')$  и  $G^E(X'')$ , так как первые всегда являются отрезками  $[\alpha', \beta']$  и  $[\alpha'', \beta'']$  соответственно. Поскольку график функции  $d_x^{E\Delta}(X)$  – ломаная с одним участком  $[\alpha, \beta]$  убывания от 1 до 0, то  $G^{E\Delta}(X') \succeq G^{E\Delta}(X'')$  верно тогда и только тогда, когда  $\alpha' \geq \alpha''$  и  $\beta' \geq \beta''$ .

#### 1.1.10 Случай данных с неопределенностями

Пусть заданы не точные данные – числа  $x_i$ , а данные с неопределенностями  $\tilde{x}_i$ . Причиной появления неопределенности могут быть ошибки измерения, округление данных, их группировки и т.д.

Если эти данные – случайные величины  $\tilde{x}_i$ , то критерии оказываются функциями случайных величин:  $\tilde{f}_i(x) = |x - \tilde{x}_i|$ . Если известны распределения вероятностей случайных величин  $\tilde{x}_i$ , то можно перейти к математическим ожи-

даниям  $\bar{f}_i(x) = E[\tilde{f}_i(x)]$ . В частности, если случайная величина  $\tilde{x}_i$  имеет равномерное распределение на отрезке  $[a_i, b_i]$  (например, вследствие округления данных), то:

$$\bar{f}_i(x) = \begin{cases} \frac{1}{2}(a_i + b_i) - x, & x \leq a_i, \\ \frac{1}{b_i - a_i} \left( x^2 - x(a_i + b_i) + \frac{1}{2}(a_i^2 + b_i^2) \right), & a_i < x < b_i, \\ x - \frac{1}{2}(a_i + b_i), & x \geq b_i. \end{cases} \quad (1.19)$$

График функции (1.19) представлен на рисунке 1.9.

Если известны только множества  $X_i$  возможных значений данных  $\tilde{x}_i$ , то можно воспользоваться одним из принципов принятия решений при неопределенности [14]. Для принципа недостаточного основания, согласно которому распределения вероятностей на  $X_i$  принимаются равномерными, получаем  $\bar{f}_i(x) = E[\tilde{f}_i(x)]$  для этих распределений. Для принципа максимина получаем  $f_i^*(x) = \max_{x_i \in X_i} |x - x_i|$ . Если множество  $X_i$  есть отрезок  $[a_i, b_i]$ , то, как легко видеть,

$$f_i^*(x) = \begin{cases} b_i - x, & x \leq \frac{1}{2}(a_i + b_i), \\ x - a_i, & x > \frac{1}{2}(a_i + b_i). \end{cases} \quad (1.20)$$

Можно использовать и общепринятый в статистике прием – заменить интервалы  $[a_i, b_i]$  их серединами – точками  $\frac{1}{2}(a_i + b_i)$ , и тогда критериями будут функции

$$f_i^c(x) = \left| x - \frac{1}{2}(a_i + b_i) \right|. \quad (1.21)$$

Графики функций (1.20) и (1.21) представлены на рисунке 1.9.

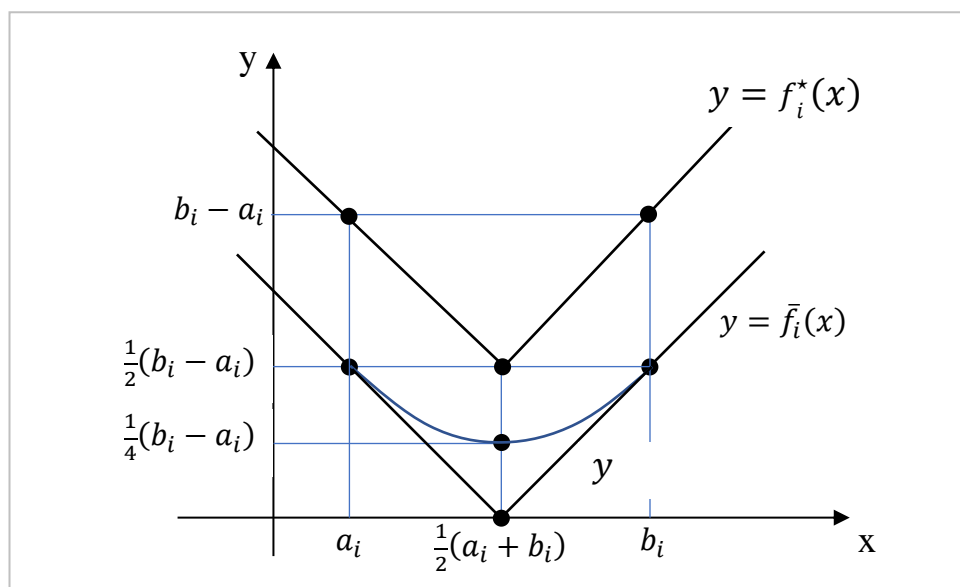


Рисунок 1.9 – Графики функций  $y = f_i^*(x)$ ,  $y = \bar{f}_i(x)$  и  $y = f_i^c(x)$  (вне отрезка  $[a_i, b_i]$  графики двух последних функций совпадают)

Таким образом, задача формально оказывается приведенной в рамки рассмотренного ранее случая точных данных  $x_i$ . Какой из возможных способов такого приведения лучше использовать при решении той или иной прикладной задачи требует специальных исследований. Однако можно сразу указать следующее:

1) Верно равенство  $f_i^*(x) = f_i^c(x) + \frac{1}{2}(b_i - a_i)$ . Поэтому если длины всех интервалов равны, то все соответствующие ПН-средние будут равными.

2) Поскольку  $\bar{f}_i(x)$  и  $f_i^c(x)$  вне отрезка  $[a_i, b_i]$  равны, то при «небольших» длинах этих отрезков по сравнению с размахом величин данных  $b_n - a_1$  соответствующие ПН-средние будут отличаться «не сильно».

#### 1.1.11 Заключение

Предложены новые понятия средних, основанные на идее многокритериальной оптимизации. Эти величины не предполагают наличия некоторых априорных свойств (что типично при классическом подходе к выбору средних в статистике и иногда может затруднять выбор среднего в конкретной задаче

[4]), но используют сведения о предпочтительности тех или иных мер удаленности между векторами – значениями векторного критерия, состоящего из расстояний от текущей точки до каждой из заданных.

Оказалось, что такие средние являются множественными, причем множества некоторых средних могут иметь достаточно сложную структуру. Это затрудняет построение множеств таких средних для выборок большого размера (при большом числе  $n$ ), но при современном уровне развития вычислительной техники и информационных технологий является эффективно преодолимым затруднением.

Предложенные средние значения представляются полезным дополнением (но, разумеется, не альтернативой) к арсеналу известных в статистике средних величин, но могут использоваться и самостоятельно.

Поставленные перед началом исследований задачи выполнены полностью. Результаты проведенных исследований нашли отражение в статье [15].

## **1.2 Разработка и анализ моделей расчета степени манипулируемости правил агрегирования предпочтений**

### **1.2.1 Введение**

В теории коллективного выбора одной из центральных тем является проблема стратегического поведения участников голосования: один или несколько избирателей могут повлиять на результат голосования в лучшую для себя сторону, намеренно исказив свои предпочтения. Если процедура агрегирования допускает возможность стратегического поведения избирателей, то она называется манипулируемой. В работах [16, 17] было показано, что не существует процедуры агрегирования, которая была бы одновременно и недиктаторской, и неманипулируемой. Чтобы лучше понять, в какой степени и при каких условиях правила коллективного выбора подвержены манипулированию, было разработано много математических моделей и выработано несколько широко используемых подходов к изучению проблемы. В рамках проекта МЦАВР 2020 года были решены задачи, которые вносят вклад в область

теоретического и алгоритмического исследования проблемы манипулирования. Так, степень манипулируемости правила в большой степени зависит от выбора математической модели и её параметров. Поэтому одна из работ этого направления посвящена созданию и описанию единого программного комплекса для расчета индексов манипулируемости. Разработанный комплекс программ охватывает большое разнообразие начальных параметров модели и позволяет эффективно вычислять степень манипулируемости для 27 различных правил.

Также была продолжена работа в направлении теоретического исследования безопасности коалиционного манипулирования, начатого в 2019 году. Поставлен следующий вопрос: при каких условиях несогласованность действий внутри коалиции манипулирующих избирателей приведет к ухудшению для них результата голосования? Пусть коалиция избирателей имеет возможность повлиять на результат голосования в лучшую для себя сторону, при условии, что манипулируют все члены коалиции. Небезопасным манипулирование называется в случае, если результат будет хуже, чем изначально, при условии, что не все члены коалиции решают манипулировать. В рамках этого направления в 2020 году были доказаны теоремы о необходимых и достаточных условиях существования небезопасного коалиционного манипулирования для правила Коупленда и двухступенчатой мажоритарной системы.

1.2.2 Алгоритмы для оценки индексов манипулируемости процедур агрегирования

1.2.2.1 Обзор литературы и постановка задачи

В целом ряде работ рассматривалась проблема манипулируемости для различных процедур агрегирования. В одной из первых работ по сравнению различных процедур агрегирования [18] были рассмотрены четыре процедуры: правило Борда, процедура Хара, правило относительного большинства и правило Кумбса. Дальнейшие работы в этой области обычно включали большее количество процедур агрегирования. Самое большое их количество – 26

процедур - можно найти в [19], где рассматривалось индивидуальное манипулирование в основном для небольшого количества участников. В [20] были рассмотрены случаи большего количества участников, но для индивидуального манипулирования и трёх альтернатив. В [21] были рассмотрены и случаи для 4 и 5 альтернатив в индивидуальном манипулировании.

Частные случаи, когда выбираются процедуры какого-то типа для какого-то количества участников или альтернатив, могут описываться аналитически. Например, в [22] можно увидеть аналитическое исследование для скоринговых процедур агрегирования для 3 альтернатив. Аналитический вывод формул можно увидеть и в [23], где рассматриваются также и разные модели манипулирования. Проблема в том, что для всех процедур агрегирования (не только скоринговых, но и мажоритарных и т.д.) для случаев с разными количествами альтернатив (не только 3, но и 4 и т.д.), а также не только для индивидуального, но и для коалиционного манипулирования вывести аналитические формулы представляется невозможным, именно поэтому для анализа большего количества процедур и всех возможных случаев мы применяем компьютерное моделирование.

Степень манипулируемости процедуры агрегирования оценивается как доля ситуаций, когда манипулирование возможно, к общему количеству ситуаций (профилей). Так определяется индекс Нитцана-Келли, который был введен в [24], [25].

Вычисление степени манипулируемости заданной процедуры агрегирования – непростая алгоритмическая задача, так как невозможно сгенерировать все возможные профили из-за их огромного количества. Уже при небольшом количестве участников голосования, например, при 20 участниках, комбинаторная сложность рассмотрения всех возможных попыток манипулирования может превышать вычислительные мощности современных ЭВМ. Поэтому используется генерация определенного количества случайных профилей, чтобы для них посчитать индексы манипулируемости.

### 1.2.2.2 Основные обозначения

Мы используем обозначения из [20]. Количество участников голосования обозначаем через  $n$ , количество альтернатив (кандидатов) –  $m$ . Каждый кандидат (альтернатива) обозначается строчной буквой латинского алфавита. У каждого участника голосования есть предпочтения на множестве альтернатив, выраженное линейным порядком, т.е. у каждого участника одно из  $m!$  возможных предпочтений.

Профиль – это совокупность участников голосования и их предпочтений на множестве кандидатов, то есть множество из  $n$  линейных порядков на множестве из  $m$  альтернатив. Процедура агрегирования – это отображение множества профилей на множество всех возможных результатов.

Для представления результата процедуры агрегирования мы используем две концепции. Обе дают ответ на вопрос, что будет, если несколько альтернатив получат одинаковое количество голосов. Первая – множественный выбор. Предполагается, что если у нескольких альтернатив одинаковое количество голосов, то все они войдут в результат процедуры агрегирования. В этом подходе всего возможно  $2^m - 1$  исходов, т.е. результат – это подмножество множества альтернатив за вычетом пустого множества. А как определить, какой выбор будет для участника лучше? Например, если до манипулирования результат процедуры агрегирования был  $\{b\}$ , а после он станет равным  $\{a, c\}$ , это будет вариант лучше или хуже? В литературе используются методы построения расширенных предпочтений, т.е. предпочтений участника не на множестве альтернатив, а на множестве всех возможных подмножеств множества альтернатив (всех возможных множественных выборов).

Для 3 альтернатив известно 4 способа построения расширенных предпочтений: Лексимин, Лексимакс, Рискофил, Рискофоб [20]. Например, в Лексимаксе при выборе между двумя результатами (подмножествами множества альтернатив) участник предпочтёт тот, в котором самая хорошая альтернатива будет как можно лучше для него. То есть в Лекси-



максе  $\{a, c\}$  будет предпочтительнее, чем  $\{b\}$ . Если лучшие альтернативы в выборе одинаковые, то сравниваются вторые наилучшие и т.д. Поэтому для Лексимаксе  $\{a, b, c\}$  будет лучше чем  $\{a, c\}$ .

Для 4 альтернатив известно 10 способов построения расширенных предпочтений, для 5 альтернатив – 12 методов. В работе мы рассматриваем подсчёт результатов для каждого метода построения расширенных предпочтений. Итак, в концепции множественного выбора процедура агрегирования – это отображение множества профилей на множество исходов, где множество исходов – это множество подмножеств множества альтернатив за вычетом пустого множества (его мощность  $2^m - 1$ ).

Вторая концепция – это алфавитное устранение несравнимости. Если несколько альтернатив набрали одинаковый результат, то побеждает та, которая меньше лексикографически. Например, если альтернативы  $a$  и  $b$  набрали по 5 голосов в процедуре относительного большинства, а другие альтернативы – строго меньше, то исходом будет  $\{a\}$ . В случае алфавитного устранения несравнимости процедура агрегирования – это отображение множества профилей на множество исходов, где множество исходов – это множество альтернатив (его мощность  $m$ ).

Манипулирование происходит, если существует участник (или группа участников), который записывают в бюллетень неискренние предпочтения, после чего выбор оказывается для него (них) лучше, чем если бы он вписал в бюллетень свои искренние предпочтения.

Задача исследования – найти, какие процедуры агрегирования являются наименее манипулируемыми для различных ситуаций.

### 1.2.2.3 Процедуры агрегирования

В литературе известно немало процедур агрегирования, и разные работы исследуют различные процедуры по отдельности или в группе из нескольких. Нашей задачей было найти наименее манипулируемые процедуры агрегирования из наибольшего возможного количества известных процедур, поэтому мы взяли за основу список процедур агрегирования из [19],[20], так как именно

в этих работах приводится один из наиболее полных списков известных процедур агрегирования.

Таким образом, первая группа процедур – это процедуры агрегирования в том виде, в котором они изложены в [19], [20]. Список процедур агрегирования, а также примерные алгоритмические оценки времени расчета результатов процедур агрегирования для одного профиля указаны в Таблице 1.1.

Таблица 1.1 – Процедуры агрегирования и асимптотическое время их расчёта

Процедура агрегирования	Асимптотическое время расчета
Правило относительного большинства	$O(n+m)$
Одобряющее голосование с $q=2$	$O(n+m)$
Правило Борда	$O(n \cdot m)$
Процедура Блэка	$O(n \cdot m^2)$
Обратное правило Борда	$O(n \cdot m^2)$
Пороговое правило	$O(m(n+m))$
Процедура Хара	$O(n \cdot m^2)$
Обратное правило относительного большинства	$O(n+m)$
Процедура Нансона	$O(n \cdot m^2)$
Процедура Кумбса	$O(n \cdot m^2)$
Минимальное доминирующее множество	$O((n+2m) \cdot m^2)$
Минимальное недоминируемое множество	$O((n+2m) \cdot m^2)$
Непокрытое множество I	$O(n \cdot m^2)$
Непокрытое множество II	$O(n \cdot m^2)$
Процедура агрегирования	Асимптотическое время расчета
Правило Ричельсона	$O(n \cdot m^2)$
Минимальное слабоустойчивое множество	$O((n+2m) \cdot m^2)$
Правило Фишберна	$O(n \cdot m^2)$
Правило Коупланда I	$O(n \cdot m^2)$
Правило Коупланда II	$O(n \cdot m^2)$
Правило Коупланда III	$O(n \cdot m^2)$
Процедура Симпсона	$O(n \cdot m^2)$
Процедура минмакс	$O(n \cdot m^2)$
Сильное q-Паретовское правило простого большинства	$O(m(n+m))$
Сильное q-Паретовское правило относительного большинства	$O(m(n+m))$
1-устойчивое множество	$O(m^2(n+m+2m))$
2-устойчивое множество	$O(m^2(n+m+2m))$
3-устойчивое множество	$O(m^2(n+m+2m))$

При реализации каждой процедуры использовались их описания из [19]. При реализации использовались оптимизации, связанные с исключением расчёта одних и тех же вспомогательных данных несколько раз. Например, если

речь идёт о мажоритарных правилах, для которых нужно высчитывать мажоритарную матрицу, то она высчитывалась для заданного профиля один раз, затем использовалась для всех мажоритарных процедур. За счёт этого достигается небольшая экономия во времени расчётов.

#### 1.2.2.4 Impartial Culture (IC) и Impartial Anonymous Culture (IAC)

При оценке степени манипулируемости процедур агрегирования возникает вопрос: каковы вероятности появления тех или иных профилей голосования, т.е. наборов множеств участников голосования, у каждого из которых есть предпочтения на множестве альтернатив?

Две наиболее популярные в литературе концепции – это Impartial Culture (IC) и Impartial Anonymous Culture (IAC).

В рамках IC предполагается, что предпочтения участников независимы и равновероятны. С точки зрения моделирования это эквивалентно генерации одного из  $m!$  различных линейных порядков в качестве предпочтений для каждого из  $n$  участников. Всего получается  $m!^n$  различных возможных профилей голосования.

В рамках модели IAC предполагается, что профили голосования равновероятны с учетом анонимности, т.е. все возможные профили делятся на классы эквивалентности, где два профиля попадают в один класс, если они могут быть получены переименованием избирателей. И в IAC предполагается, что каждый класс эквивалентности равновероятен. Общее количество различных профилей в IAC равно  $C_{m!+n-1}^n$ .

При оценке степени манипулируемости процедур агрегирования мы рассматриваем для каждого правила оба случая: и IC, и IAC. Из результатов видно, что степени манипулируемости для правил в IC и IAC различаются, например, индексы манипулируемости могут отличаться в два раза, хотя есть и много сходств по тому, какие процедуры оказываются наименее манипулируемыми для разных ситуаций.

Общее количество профилей для случаев IC и IAC очень велико ( $m!^n$  и  $C_{m!+n-1}^n$ , соответственно), и за разумное время сгенерировать и определить манипулируемость всех возможных профилей невозможно.

Поэтому мы генерируем 1 миллион профилей для каждой возможной ситуации, после чего рассчитываем коэффициенты манипулируемости для данных профилей. В [26] показано, что при оценке индекса манипулируемости при случайной генерации 1 миллиона профилей погрешность при вычислении индекса Нитцана-Келли не превышает 0.001. Такая погрешность даёт возможность оценить, какие процедуры агрегирования являются наименее манипулируемыми для различных ситуаций.

#### 1.2.2.5 Индивидуальное и коалиционное манипулирование

Какое количество участников голосования может пытаться манипулировать? Если только один участник предъявляет неискренние предпочтения, то это будет индивидуальное манипулирование. Если группа участников объединяется, то это коалиционное манипулирование.

Важно подчеркнуть, что профиль является манипулируемым, если существует хотя бы один способ для какого-то участника (в случае индивидуального манипулирования) или у группы участников (случай коалиционного манипулирования), при котором предъявление неискренних предпочтений повлечет за собой более выгодный результат голосования. Отсюда следует, что если мы проверяем профиль на манипулируемость, то мы должны рассмотреть все возможные попытки манипулирования. И если мы находим хотя бы одну успешную возможную попытку, мы можем сразу пометить профиль как манипулируемый (для заданной процедуры агрегирования и рассматриваемого случая). Но если рассмотренная попытка не была успешной, мы должны продолжать рассматривать все возможные попытки.

В случае с индивидуальным манипулированием надо рассмотреть всех участников голосования и все возможные способы предъявления неискренних предпочтений для каждого участника голосования. Для случая  $m$  альтернатив

получится  $m! - 1$  возможных попыток манипулирования для каждого участника. Если всего участников  $n$ , то всего получится  $n \cdot (m! - 1)$  попыток для всех участников, которые мы должны перебрать.

Здесь возможно провести оптимизацию перебора, которая будет особенно эффективна для случаев небольшого количества альтернатив и большого количества участников. Например, для случая  $n=100$  и  $m=3$ , мы будем перебирать  $100 \cdot (3! - 1) = 500$  попыток манипулирования. Но если вспомнить предпосылку, что для нас все участники голосования эквивалентны, то очевидно, что нет смысла перебирать попытки манипулирования для двух участников голосования, если у них абсолютно одинаковые предпочтения (одинаковые линейные порядки на множестве альтернатив). Например, если для какого-то участника  $a > b > c$  и мы сгенерировали все возможные попытки манипулирования, то мы можем не перебирать все возможные попытки манипулирования для других участников, у которых такие же предпочтения. Это означает, что максимальное количество участников, для которых мы будем перебирать все возможные попытки манипулирования, ограничено числом  $m!$ , то есть сложность перебора будет равна  $(m! - 1) \cdot \min(n, m!)$ . И для случая  $n=100$  и  $m=3$  это будет равно не 500 попыткам (как в полном переборе), а только 30.

В случае коалиционного манипулирования ключевым аспектом являются принципы, по которым участники могут объединяться в коалиции. Здесь должны быть точные определения, участники с какими типами предпочтений могут объединяться, есть ли ограничения на размер возможных коалиций: могут ли члены коалиции предъявлять разные неискренние предпочтения или только одинаковые и т.д.

Если никаких принципов нет, то есть любой участник может объединяться с любыми в коалиции, то это будет самый тяжелый с вычислительной точки зрения случай, так как перебор по количеству коалиций может достигать  $2^n - 1$ , т.е. все возможные подмножества участников, исключая пустую коалицию. Однако в реальной ситуации объединение участников, вероятно, должно происходить вокруг чего-то общего: общих искренних предпочтений и/или

общих желаний «продвинуть» какую-то альтернативу с помощью неискренних предпочтений.

Мы будем рассматривать два принципа формирования коалиций. Первый принцип – это когда в коалиции могут объединяться те и только те участники, у которых одинаковые искренние предпочтения. Кроме того, при голосовании все они могут предъявить только одинаковые неискренние предпочтения. Дополнительно могут вводиться ограничения на максимальный размер коалиции. Второй принцип – это объединение в коалиции вокруг продвижения какого-то единого кандидата. Здесь будут объединяться те участники, которые хотят продвинуть единую альтернативу, при этом их предпочтения могут быть разными.

В первом случае оптимизация вычислений может быть достигнута за счёт того, что нам не нужно рассматривать все подмножества участников, достаточно рассматривать отдельно группы участников с одинаковыми предпочтениями, а уже внутри этой группы смотреть возможные коалиции. Например, если участников  $n$ , а участников с предпочтениями  $a > b > c$  в профиле ровно  $n_1$ , то мы можем рассматривать не  $2^n - 1$  коалиций, а  $2^{n_1} - 1$ . Затем, если у  $n_2$  участников предпочтения  $a > c > b$ , то для них будет  $2^{n_1} - 1$  возможных коалиций.

Однако и здесь будут получаться весьма большие размеры вычислений, поэтому можно использовать следующее наблюдение. Для нас неважно, какие именно участники будут объединяться в коалиции среди группы с одинаковыми предпочтениями, для нас важно их количество. Например, если у пяти участников предпочтения  $a > b > c$ , то случаи, когда первый и второй объединятся, будут эквивалентны случаю, когда третий и четвертый объединятся. Тогда размер перебора внутри группы участников с одинаковыми предпочтениями сводится не к перебору по подмножествам, а к перебору размера коалиции. Если  $n_1$  участников имеют предпочтения  $a > b > c$ , то мы будем перебирать  $n_1$  возможных коалиций, в каждую из которых будет входить от 1 до  $n_1$  участников. Тогда получится, что общий размер перебора сводится к

$n_1+n_2+\dots+n_6=n$ . Для случая 4-х и большего количества альтернатив также получится перебор размера  $n$ , так как будет перебираться каждая из  $m!$  групп участников с одинаковыми предпочтениями от 1 до мощности этого множества, а суммарная мощность всех множеств будет равна количеству участников голосования. Таким образом, от экспоненциальной сложности перебора коалиций мы пришли к линейной сложности.

Второй способ формирования коалиций – это предположение, что участники коалиций объединяются, чтобы вместе продвинуть некоего кандидата. Это более гибкая предпосылка по сравнению с первым методом формирования коалиций: здесь объединяться могут участники с разными предпочтениями, а также записывать в бюллетени они могут разные искренние или неискренние предпочтения. Из-за высокой комбинаторной сложности для данного типа формирования коалиций мы рассматриваем только случай трёх альтернатив.

В данном случае вновь мы могли бы иметь экспоненциальную сложность перебора, однако она превращается в квадратичную для случая трёх альтернатив. По определению этого типа коалиционного манипулирования, из  $m! = 3! = 6$  групп участников только две могут объединяться, чтобы продвинуть единую альтернативу. Тогда будет два цикла перебора: от 0 до  $n_1$  (сколько участников с предпочтениями первого типа входят в коалицию) и от 0 до  $n_2$  (сколько участников с предпочтениями второго типа входят в коалицию). В итоге – квадратичная сложность вместо экспоненциальной.

#### 1.2.2.6 Коэффициенты манипулируемости

В исследованиях по манипулируемости используется целый ряд различных коэффициентов. Перечислим те из них, которые рассматриваются ниже.

Один из самых популярных индексов манипулируемости — это индекс Нитцана-Келли (НК), который был впервые введён в [24], [25]. Индекс равен доле манипулируемых профилей в общем количестве профилей. Формула выглядит следующим образом (1.22):

$$NK = \frac{d_0}{d_{\text{общее}}}, \quad (1.22)$$

где  $d_0$  – количество манипулируемых профилей, т.е. профилей, где существует хотя бы один успешный способ манипулирования, а  $d_{\text{общее}}$  – общее количество профилей.

В случае с генерацией 1 миллиона профилей в знаменателе будет находиться 1 миллион, а в числителе – количество профилей из сгенерированного 1 миллиона, для которых существует хотя бы одна успешная попытка манипулирования.

Помимо индекса НК в исследованиях встречались и другие индексы. Например, в [19] был исследован индекс свободы манипулирования. Рассмотрим все возможные попытки манипулирования, которые может предпринять участник голосования. Их количество равно  $m! - 1$ . При каждой попытке манипулирования после искажения предпочтений может реализоваться один из трёх случаев:

- выбор стал лучше для манипулирующего участника;
- выбор остался таким же для манипулирующего участника;
- выбор стал хуже для манипулирующего участника.

Тогда были предложены три коэффициента свободы манипулирования:  $I_1^+$ ,  $I_1^0$  и  $I_1^-$ , где каждый будет равен доле случаев, которые приведут к улучшению, неизменности или ухудшению выбора. Иначе говоря, если задан профиль, то для каждой попытки манипулирования смотрим, каким будет выбор. Если лучше, чем начальный, то  $I_1^+$  увеличивается на 1. Если эквивалентный начальному, то  $I_1^0$  увеличивается на 1. Если хуже начального, то  $I_1^-$  увеличивается на 1. Индексы нормируются на количество попыток манипулирования, т.е. каждый делится на  $m! - 1$ . В итоге получится, что  $I_1^+ + I_1^0 + I_1^- = 1$ .

Помимо свободы манипулирования в [19] был введен индекс эффективности манипулирования. Рассмотрим все попытки предъявить неискренние предпочтения ( $m! - 1$ ), и из них возьмём те, которые приводят к улучшению



результата процедуры агрегирования для манипулирующего участника. В каждой такой попытке мы можем рассчитать, на какое количество мест в его расширенных предпочтениях улучшится итоговый выбор. Возьмём среднее значение по каждой успешной возможной попытке манипулирования. Усредним эти значения по всем участникам в профиле, затем усредним по всем профилям. В итоге получится индекс  $I_2$ , являющийся индексом эффективности манипулирования.

Есть вторая вариация индекса эффективности манипулирования – индекс  $I_3$ . В этом случае мы будем не усреднять все возможные выигрыши, измеряемые в количестве мест продвижения в предпочтениях участника, а брать максимальное. Иначе говоря, схема выглядит следующим образом:

- 1) выбираем профиль;
- 2) для каждого участника рассматриваем все попытки манипулирования;
- 3) выбираем максимальное улучшение в терминах количества мест, на которое может улучшиться выбор в расширенных предпочтениях участника при манипулировании;
- 4) усредняем результаты по всем участникам внутри профиля;
- 5) усредняем результаты по всем профилям.

Итак, в данной работе мы рассчитываем индексы  $NK$ ,  $I_1^+$ ,  $I_1^0$ ,  $I_1^-$ ,  $I_2$  и  $I_3$ .

#### 1.2.2.7 Алгоритмическая схема оценки степени манипулируемости

Алгоритм расчета индексов манипулируемости состоит из нескольких этапов. На первом этапе задаются параметры модели:

- 1) IAC или IC;
- 2) количество альтернатив, от 3 до 5;
- 3) количество участников голосования (от 3 до 100);
- 4) индивидуальное или коалиционное манипулирование (если коалиционное, то какого из двух описанных выше типов).

Эти параметры подаются на вход программе на следующий этап. Для оценки процедуры агрегирования на предмет манипулируемости требуется рассмотреть случаи от 3 до 100 участников голосования, для случаев IAC и IC.

Дальнейшая стадия – это генерация профилей. Для каждого профиля рассматриваются следующие ситуации и наборы параметров:

- алфавитное устранение несравнимости и каждый возможный тип построения расширенных предпочтений (3 способа для 3 альтернатив, 10 способов для 4 альтернатив и 12 способов для 5 альтернатив) для случая множественного выбора;

- каждая процедура агрегирования из списка рассматриваемых процедур агрегирования.

Для каждого возможного набора параметров проверяются все возможные попытки манипулирования.

Если речь идёт об индивидуальном манипулировании, то рассматриваются все возможные способы предъявления неискренних предпочтений одним участником. Это перебор по  $n$  участникам (но как было указано выше, не более чем по  $m!$  участникам), у каждого из которых рассматриваются  $m!-1$  попыток предъявить неискренние предпочтения.

Для случая коалиционного манипулирования рассматриваются два различных варианта в зависимости от типа коалиционного манипулирования. В первом типе делаем перебор величины  $n$ , как описывалось в пункте про коалиционное манипулирование.

Для случая коалиций, которые продвигают одного кандидата, мы ограничиваемся расчетами по 3 альтернативам, так как здесь сложность перебора вырастает существенно: нам нужно рассмотреть все возможные комбинации участников с разными предпочтениям. Однако здесь мы используем прежнюю оптимизацию, основанную на том, что для нас участники равноценны. Так, если  $n_1$  участников имеют предпочтения  $a > b > c$ , а  $n_2$  участников имеют предпочтения  $a > c > b$ , и требуется рассмотреть все возможные коалиции из

$k$  участников, которые состояли бы только из участников с такими предпочтениями, то нам не нужно делать полный перебор размера  $C_{n_1+n_2}^k$ , нам достаточно провести перебор размера  $k + 1$ , так как мы будем перебирать, сколько (от 0 до  $k$  включительно) из участников будут из первой группы (остальные из второй). Тот факт, что мы избегаем количества сочетаний и сложной комбинаторной сложности в рассмотрении коалиций, позволяет сделать оценку степени коалиционной манипулируемости за разумное время.

Рассматривая все возможные попытки манипулирования, мы делаем расчет коэффициентов манипулируемости. При этом в силу особенности коэффициентов свободы и эффективности манипулирования надо для каждого профиля рассмотреть все возможные попытки манипулирования (а не ограничиться нахождением хотя бы одной успешной), чтобы рассчитать точные коэффициенты. После того как рассмотрены все попытки по всем профилям, рассчитываются значения индексов манипулируемости для сгенерированных профилей и получаем итоговые значения.

Очевидно, что вычислительная сложность схемы манипулирования огромна, так как сначала генерируется около тысячи комбинаций параметров (разные количества участников, альтернатив, рассматриваются различные методы построения расширенных предпочтений), затем генерируется миллион профилей, после чего следует генерация всех возможных попыток манипулирования.

Примерная сложность описанной схемы составляет:

$$\begin{aligned} \sum_{m=3,4,5} \sum_{n=3}^{100} \sum_{k=1}^n 4 \cdot 27 \cdot 1\,000\,000 \cdot C_n^k \cdot k \cdot m \cdot (m! - 1) \cdot n \\ = \sum_{m=3,4,5} \sum_{n=3}^{100} \sum_{k=1}^n 1.08 \cdot 10^8 \cdot C_n^k \cdot k \cdot n \cdot m \cdot (m! - 1) \end{aligned}$$

операций, что для случая с 5 альтернативами ( $m=5$ ) и 100 участниками голосования ( $n=100$ ) делает невозможным полные расчеты на современных суперкомпьютерах даже за несколько сотен лет.

Именно поэтому важны описанные выше методики оптимизации, особенно для случая коалиционной манипулируемости. Благодаря эквивалентности участников с одинаковыми предпочтениями мы не рассматриваем все возможные коалиции (чисто сочетаний), а рассматриваем по одному случаю из множества эквивалентных ситуаций.

После применения этих и нескольких других алгоритмических оптимизаций мы добились следующего времени работы алгоритма (в операциях).

Итоговая сложность вычислений равна

$$\begin{aligned} \sum_{m=3,4,5} \sum_{n=3}^{100} \sum_{k=1}^n 4 \cdot 1\,000\,000 \cdot opt \cdot n \cdot m! \cdot \sum_{i=1}^{27} O_i(C(\vec{P})) \\ = \sum_{m=3,4,5} \sum_{n=3}^{100} \sum_{k=1}^n 4 \cdot 10^6 \cdot opt \cdot n \cdot m! \cdot \sum_{i=1}^{27} O_i(C(\vec{P})) \end{aligned}$$

Такая сложность позволяет получить оценку манипулируемости 27 правил коллективного выбора для случая 5 альтернатив и 100 участников голосования всего лишь за несколько недель.

#### 1.2.2.8 Архитектура программного обеспечения

В рамках проекта по оценке степени манипулируемости процедур агрегирования было разработано программное обеспечение, которое позволяет рассчитывать различные индексы манипулируемости для различных входных параметров.

Ключевые параметры системы:

1) Гибкость к добавлению новых процедур агрегирования. Были получены результаты по наименее манипулируемым процедурам. Но если потребуется оценить манипулируемость какой-то новой процедуры, то текущее ПО позволит это быстро сделать.

2) Гибкость в настройке параметров. Помимо различных процедур агрегирования есть и другие параметры, которые могут меняться. Например,

могут добавляться новые виды построения расширенных предпочтений. Изменение параметров запуска программы становится очень простым и позволяет в короткие сроки проводить новые исследования.

3) Возможность задавать параметры в конфигурационных файлах и/или в интерфейсе программы, чтобы непрограммист мог осуществлять запуск расчетов при требуемых ему параметрах.

4) Возможность запуска расчетов на нескольких ЭВМ. Запуски программы проводились на нескольких компьютерах (в среднем на пяти) одновременно для ускорения расчетов. В программе должны быть предусмотрены возможности параллельных запусков и сбора расчетов из нескольких пакетов вычислений в единый файл.

5) Устойчивость программного обеспечения к сбою операционной системы, отключению электричества и т.д. Для большого количества участников голосования (например, для  $n=100$ ,  $m=5$ ) задача вычисления индексов манипулируемости может занимать несколько недель, даже на нескольких компьютерах. В случае внезапного сбоя на компьютере должна быть возможность перезапустить расчет с минимальными потерями.

В качестве языка программирования был выбран C# – высокоуровневый язык программирования с возможностью создавать удобные интерфейсы и разрабатывать сложные методы для расчетов индексов манипулируемости.

#### 1.2.2.9 Время работы алгоритмов

В Таблице 1.2 представлено примерное время работы программного обеспечения для расчета индексов манипулируемости для различных значений параметров.

Таблица 1.2 – Время работы алгоритмов для различных случаев

Количество альтернатив (m)	Количество участников (n)	Время (в часах), требуемое для расчетов индексов манипулируемости на одной ЭВМ
3	10	2
3	20	5
3	50	19

## Продолжение таблицы 1.2

3	80	42
3	100	63
4	10	13
4	20	33
4	50	138
4	80	311
4	100	467

Если мы возьмём один из самых времязатратных случаев, то есть случай  $n = 100, m = 4$  и разобьём вычисления на несколько параллельно работающих ЭВМ, то процесс займёт всего лишь несколько дней, когда для случая полного перебора потребовалось бы несоизмеримо большее время (для случая  $n = 100, m = 4$  полный перебор потребовал бы примерно  $6.3 \cdot 10^{43}$  операций для оценки степени манипулируемости 27 процедур агрегирования). В итоге все вычисления были проделаны суммарно за несколько месяцев работы нескольких ЭВМ в ИПУ РАН.

### 1.2.2.10 Проверка результатов

Для проверки корректности работы алгоритмов мы сравнили полученные результаты с известными в литературе. Чаще всего, это какие-то частные случаи, которые были подсчитаны в предыдущих работах на тему манипулируемости.

Для индивидуального манипулирования проверка результатов была произведена с работами [19] – [21]. Для коалиционного манипулирования проверка результатов проводилась по частным случаям, так как индивидуальное манипулирование – это случай, когда коалиция состоит ровно из 1 участника.

Проведенные проверки показали, что полученные результаты с точностью до погрешности сошлись с другими известными в литературе.

### 1.2.2.11 Примеры результатов

Рассчитанные индексы степени манипулируемости процедур агрегирования были опубликованы в нескольких работах. Для каждого набора входных параметров (количество альтернатив, метод построения расширенных предпо-

чений, тип манипулирования) мы можем построить график различных индексов манипулируемости, рассмотренных выше. Так как всего комбинаций параметров очень много, то все возможные графики привести в одной статье невозможно. Поэтому мы ограничимся примером графика для случая индивидуального манипулирования для 3 альтернатив, метод построения расширенных предпочтений Лексимакс, IAC. Приводятся значения индекса Нитцана-Келли.

На рисунке 1.10 мы видим, что процедура Хара является наименее манипулируемой для этого случая. Индекс НК для неё находится в пределах 0.1-0.15, что означает, что от 10 до 15% профилей являются манипулируемыми. Одобряющее голосование, когда выбираются ровно две кандидатуры от каждого участника, наоборот, является одним из самых манипулируемых правил.

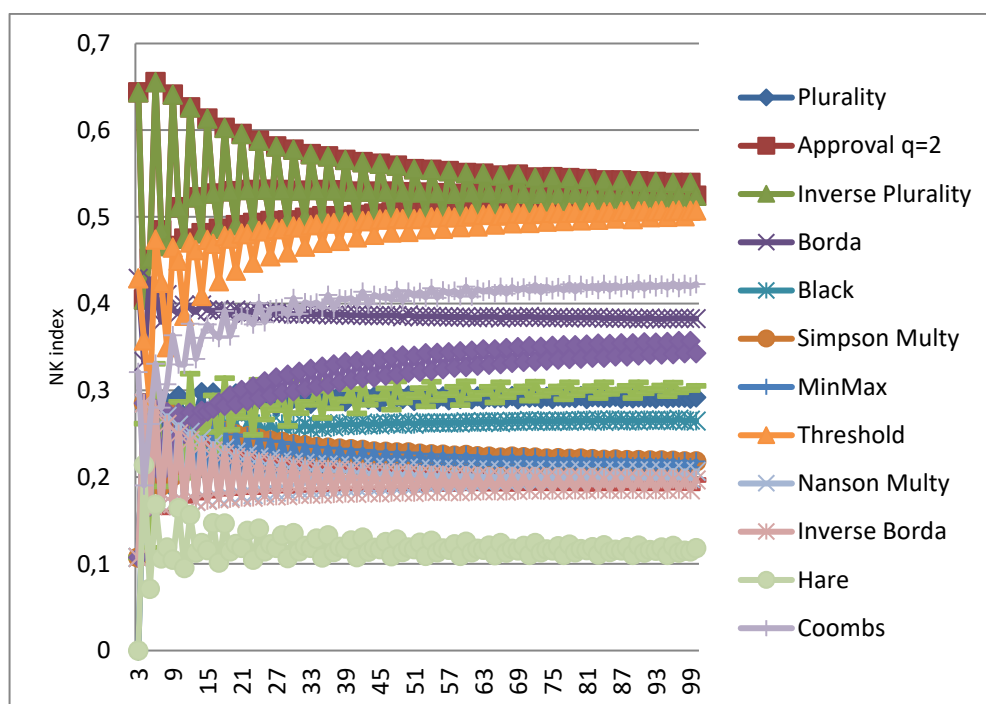


Рисунок 1.10 – Индекс НК для случая: Лексимакс, 3 альтернативы, IAC

Из результатов для разных случаев мы можем увидеть, что не существует единой процедуры агрегирования, которая была бы наименее манипулируемой для всех случаев, т.е. комбинаций параметров. Для различных случаев разные процедуры являются наименее манипулируемыми. Однако отметим, что чаще других наименее манипулируемыми оказываются процедура Хара (как видно на графике выше) и правило Нансона.

### 1.2.3 Безопасное коалиционное манипулирование

#### 1.2.3.1 Постановка задачи

Чем больше группа избирателей, тем большие она имеет возможности для влияния на результат голосования, если действует стратегически. Однако тогда существует проблема согласования действий группы, так как нескоординированные действия могут привести к результату худшему, чем был изначально. В данной работе рассматривается наиболее простая постановка задачи: пусть избиратели, имеющие одинаковые предпочтения, составляют коалицию. Предполагается, что коалиция имеет стимул манипулировать, если существует такая стратегия (неискреннее предпочтение), что результат голосования для членов коалиции будет лучше, при условии, что все члены коалиции применяют эту стратегию. Если же существует возможность получить результат худший, чем был изначально, при условии, что не все члены коалиции манипулируют, то коалиционное манипулирование небезопасно. В проекте МЛАВР 2019 года были исследованы условия существования небезопасного коалиционного манипулирования для правил подсчета очков [1],[27]. В рамках проекта МЦАВР 2020 года были найдены необходимые и достаточные условия существования небезопасно манипулируемого профиля (количество избирателей и количество альтернатив, при которых среди всех возможных профилей найдется тот, в котором существует небезопасное коалиционное манипулирование) для двухступенчатой мажоритарной системы и правила Коупленда.

#### 1.2.3.2 Описание модели

##### 1.2.3.2.1 Термины и обозначения

Обозначим за  $m$  число альтернатив, за  $n$  – количество избирателей.

1) Пусть имеется множество из  $n$  избирателей,  $N = \{1, \dots, n\}$ , каждый из которых имеет предпочтения на множестве  $X$ , состоящем из  $m$  альтернатив,  $|X| = m$ . Предпочтения избирателей представлены линейным порядком на  $X$ ,  $L(X)$  – множество всех линейных порядков. Предпочтения избирателя  $i$



обозначим за  $P_i$ , если альтернатива  $a$  более предпочтительна, чем альтернатива  $b$  для избирателя  $i$ , пишем  $aP_ib$ . Профиль предпочтений  $n$  избирателей обозначим за  $\vec{P} = (P_1, \dots, P_n) \in L(X)^N$ .

Сужением профиля на множество  $A \subseteq X$  называется  $\vec{P}/A = (P_1/A, \dots, P_n/A)$ , где  $P_i/A = P_i \cap (A \times A)$ . Пусть  $v_j(a, \vec{P})$  – количество избирателей, у которых альтернатива  $a$  занимает позицию с номером  $j$  в предпочтениях (наиболее предпочтительная альтернатива занимает первое место), а  $v(a, \vec{P}) = (v_1(a, \vec{P}), \dots, v_m(a, \vec{P}))$  – вектор позиций для альтернативы  $a$ .

2) Матрица взвешенного графа мажоритарного отношения для профиля  $\vec{P}$  обозначается за  $WMG(\vec{P})$  и состоит из элементов:  $WMG(\vec{P})_{kl} = |\{i \in N: a_k P_i a_l\}|$ . За  $\mu$  обозначим мажоритарное отношение:  $a_k \mu a_l$ , если  $WMG(\vec{P})_{kl} > WMG(\vec{P})_{lk}$ .

Матрица графа мажоритарного отношения –  $MG(\vec{P})$ , элементы которой

$$MG(\vec{P})_{kl} = \begin{cases} 1, & \text{если } a_k \mu a_l, \\ -1, & \text{если } a_l \mu a_k, \\ 0, & \text{иначе.} \end{cases}$$

Правило коллективного выбора  $C: L(X)^N \rightarrow 2^X \setminus \{\emptyset\}$  по заданному профилю предпочтений определяет множество альтернатив, являющихся коллективным выбором. Описание рассматриваемых в исследовании правил дано в следующем подразделе.

Результат правила коллективного выбора может состоять из нескольких альтернатив, поэтому для устранения множественности выбора используется разрешающее правило (tie-breaking rule), основанное на алфавитном порядке: предположим фиксированным некоторый линейный порядок  $P_T$  на  $X$ ,  $aP_T bP_T cP_T \dots$ . Если результат правила коллективного выбора состоит из нескольких альтернатив, то среди них выбирается альтернатива, недоминируемая по  $P_T$ . Применение разрешающего правила возможно и в качестве функции:  $T(A) = \{a \in A: \forall x \in A, x \neq a, (a, x) \in P_T\}$ .

Рассмотрим избирателя  $i$ . Пусть  $\vec{P}$  – множество избирателей, имеющих предпочтения, идентичные предпочтениям избирателя  $i$ , обозначается за  $K(i, \vec{P})$  – коалиция, в которую входит  $i$ . Профиль предпочтения членов коалиции обозначим за  $\vec{P}_K$ , а профиль предпочтений всех избирателей, кроме членов коалиции – за  $\vec{P}_{-K}$ . Таким образом, профиль предпочтений  $\vec{P}$  может быть также записан как  $(\vec{P}_K, \vec{P}_{-K})$ . Пусть  $\tilde{P} \neq P_i$  – неискренние предпочтения. Профиль предпочтений, в котором члены коалиции изменили предпочтения на  $\tilde{P}$ , обозначается за  $(\tilde{P}_K, \vec{P}_{-K})$ .

Определение 1.1. Пусть дано правило коллективного выбора  $C$ , профиль предпочтений  $\vec{P}$  и коалиция  $K$ . Коалиция  $K$  имеет стимул манипулировать в  $\vec{P}$  при  $C$ , если существует  $\tilde{P} \in L(X)$ , т.ч.  $T\left(C\left(\tilde{P}_K, \vec{P}_{-K}\right)\right) P_i T\left(C\left(\vec{P}\right)\right)$ .

Определение 1.2. Профиль предпочтений  $\vec{P}$  называется *коалиционно манипулируемым* при правиле  $C$ , если существует коалиция  $K$ , которая имеет стимул манипулировать в  $\vec{P}$  при  $C$ .

Определение 1.3. Правило коллективного выбора  $C$  называется *коалиционно манипулируемым*, если для существует профиль предпочтений  $\vec{P}$ , коалиционно манипулируемый при правиле  $C$ .

#### 1.2.3.2.2 Безопасное и небезопасное коалиционное манипулирование

Предположим теперь, что не все члены коалиции манипулируют – некоторая часть коалиции остается при своих искренних предпочтениях. Таковую ситуацию назовем *частичным коалиционным манипулированием*.

Определение 1.4. Пусть коалиция  $K$  имеет стимул манипулировать в  $\vec{P}$  при  $C$  посредством стратегии  $\tilde{P}$ . Пусть  $M \subset K$ ,  $M \neq \emptyset$ , и избиратели из  $M$  манипулируют, а избиратели из  $K \setminus M$  не манипулируют. Ситуация  $(\tilde{P}_M, \vec{P}_{-M})$  называется *частичным манипулированием со стороны коалиции  $K$* .

Определение 1.5. Пусть коалиция  $K$  имеет стимул манипулировать в  $\vec{P}$  при  $C$  посредством стратегии  $\tilde{P}$ . Если существует частичное манипулирование, при котором результат по правилу  $C$  хуже, чем результат при искренних предпочтениях, т.е.  $\exists M \subset K, M \neq \emptyset \ T \left( C(\vec{P}) \right) P_i \ T \left( C \left( \tilde{\vec{P}}_M, \vec{P}_{-M} \right) \right)$ , то коалиционное манипулирование *небезопасно*. Если не существует такого частичного манипулирования, то коалиционное манипулирование *безопасно*.

Другими словами, коалиционное манипулирование безопасно, если нет риска получить менее выгодных исход в случае, если кто-то из коалиции не будет манипулировать. Теперь дадим определение безопасно (небезопасно) манипулируемому профилю и правилу, аналогично Определениям 1.2 и 1.3.

Определение 1.6. Профиль предпочтений  $\vec{P}$  называется *безопасно (небезопасно) манипулируемым* при правиле  $C$ , если существует коалиция  $K$ , которая имеет стимул манипулировать в  $\vec{P}$  при  $C$  и это манипулирование безопасно (небезопасно).

Таким образом, один и тот же профиль может быть и безопасно, и небезопасно манипулируемым. Более того, он может быть безопасно и небезопасно манипулируемым одной и той же коалицией (с использованием различных стратегий).

Определение 1.7. Правило коллективного выбора  $C$  называется *безопасно (небезопасно) манипулируемым*, если для существует профиль предпочтений  $\vec{P}$ , безопасно (небезопасно) манипулируемый при правиле  $C$ .

#### 1.2.3.2.3 Правила коллективного выбора

Двухступенчатая мажоритарная система. Состоит из двух этапов:

1) Для каждой альтернативы подсчитывается количество очков по правилу относительного большинства. Вектор очков для первого этапа

$$S^1(\vec{P}) = \left( S^1(a_1, \vec{P}), \dots, S^1(a_m, \vec{P}) \right),$$

2) где  $S^1(a_j, \vec{P}) = (1, 0, \dots, 0) \cdot v(a_j, \vec{P})$ .

3) 2. Выбираются две альтернативы с наибольшим количеством очков. Если выбирается более двух альтернатив, то применяется разрешающее правило  $T$ .

$$a_k = T\left(\operatorname{argmax}_{a_j \in X} \left(S^1(a_j, \vec{P})\right)\right), a_l = T\left(\operatorname{argmax}_{a_j \in X \setminus \{a_k\}} \left(S^1(a_j, \vec{P})\right)\right).$$

В итоговый выбор входит альтернатива, доминирующая другую по мажоритарному отношению (или обе альтернативы, если голоса избирателей делятся между ними поровну:

$$a_k \in C(\vec{P}) \Leftrightarrow [WMG(\vec{P})_{kl} \geq WMG(\vec{P})_{lk}].$$

Правило Коупленда. Вычисляется граф мажоритарного отношения. Затем очки альтернатив рассчитываются следующим образом

$$CS(a_k, \vec{P}) = \sum_{l=1}^m MG(\vec{P})_{kl}.$$

Альтернатива с наибольшим количеством очков выигрывает.

$$c \in C(\vec{P}) \Leftrightarrow [\forall a \in X, CS(c, \vec{P}) \geq CS(a, \vec{P})].$$

### 1.2.3.3 Результаты

Теперь определим, при каких значениях количества избирателей и альтернатив существует небезопасно манипулируемый профиль для двухступенчатой мажоритарной системы и правила Коупленда.

Теорема 1.10. Для двухступенчатой мажоритарной системы, при  $m = 3$  не существует небезопасно манипулируемых профилей; при  $m = 4$  небезопасно манипулируемый профиль существует, если и только если  $n \geq 6$ ; при  $m \geq 5$  небезопасно манипулируемый профиль существует, если и только если  $n \geq 5$ .

Таким образом, по Теореме 1.10, двухступенчатая мажоритарная система является небезопасно манипулируемым правилом. Более подробно, любое коалиционное манипулирование безопасно для  $m = 3$ , а также для  $n =$

3, 4 и в случае  $m = 4, n = 5$ . В остальных случаях существует небезопасно манипулируемый профиль.

Теорема 1.11. Для правила Коупленда небезопасно манипулируемый профиль существует, если и только если  $m \geq 4$  и  $n \geq 4$ .

#### 1.2.4 Заключение

Мы рассмотрели подробные схемы расчета нескольких коэффициентов манипулируемости для различных ситуаций. При экспоненциальном общем количестве всех возможных профилей голосования полученная модель позволяет подсчитывать коэффициенты манипулируемости за разумное время с определенной точностью.

Мы рассмотрели большое количество различных ситуаций: случаи Impartial Culture и Impartial Anonymous Culture, различные способы построения расширенных предпочтений, различные количества альтернатив (3, 4 и 5 альтернатив), случаи от 3 до 100 участников голосования. Для каждой комбинации параметров были произведены расчеты, показывающие, какая процедура агрегирования будет наименее манипулируемой для данного случая. В качестве основного индекса манипулируемости мы использовали индекс Нитцана-Келли, но также рассмотрели расчет индексов свободы и эффективности манипулирования. Разработанные алгоритмы и программное обеспечение позволяют получать результаты для большего количества различных комбинаций параметров, чем встречалось в отдельных предыдущих исследованиях. Чаще других наименее манипулируемыми были процедуры Хара и правило Нансона, однако то, какая процедура будет являться наименее манипулируемой, зависит от параметров (количество участников голосования, количество альтернатив, метод построения расширенных предпочтений и т.д.).

Описанная схема является гибкой к добавлению новых процедур агрегирования. Несмотря на проведенные расчеты для 27 известных процедур агрегирования и обобщённых скоринговых правил, даже если потребуется оценить степень манипулируемости какой-то новой процедуры агрегирования,

это можно будет сделать с помощью описанной схемы. Схема реализована в виде программного обеспечения на языке C#.

В исследовании безопасности коалиционного манипулирования найдены условия, когда манипулирование всегда безопасно и когда существует небезопасно манипулируемый профиль для двухступенчатой мажоритарной системы и для правила Коупленда. Эти результаты могут быть использованы в дальнейшем при изучении манипулирования в условиях неполной информации: если существует неопределенность в отношении действий других участников голосования, то уменьшаются стимулы к стратегическому поведению.

Результаты исследований отражены в работах [28], [29].

### **1.3 Нахождение вероятностей исследование однопиковых предпочтений**

#### **1.3.1 Введение**

Задача агрегирования предпочтений возникает не только в теории голосования, но и во множестве других приложений: рекомендательные системы, мультиагентные системы, агрегирование экспертных оценок и т.д. В классической (Эрроувской) постановке задачи предпочтения агентов представимы бинарными отношениями, являющимися линейными порядками. Одной из основных теоретических проблем данной задачи является парадокс Кондорсе: предпочтения коллектива, построенные на основе правила простого большинства не являются транзитивным. Данный парадокс возникает при трех и более альтернативах. Теорема о невозможности Эрроу, непосредственно связанная с этим парадоксом [30], утверждает, что при трех и более альтернативах не существует аксиоматически обоснованного правила агрегирования предпочтений.

Решением теоремы о невозможности Эрроу является сужение области применения правил агрегирования. На области профилей предпочтений, для которых правило простого большинства не создает циклов в коллективных

предпочтениях, любое правило, совпадающее с правилом простого большинства в данной области, удовлетворяет условиям теоремы Эрроу.

Правило простого большинства имеет аксиоматические обоснования [31] - [34], несложно с вычислительной точки зрения и интуитивно понятно участникам голосования.

Необходимость сужения области рассматриваемых профилей предпочтений привело к теории доменов предпочтений. Под доменом предпочтений понимается множество линейных порядков. Домен генерирует профили предпочтений, состоящие только из линейных порядков, составляющих домен. Инада [35], [36] и Сен [37] описали домены предпочтений, исключающие парадокс Кондорсе. Для нечетного количества агентов правило простого большинства приводит к единственному согласованному коллективному решению. Инада и Сен выделили домены однопиковых предпочтений, однолунковых предпочтений, делимых предпочтений.

Однопиковые предпочтения предполагают некоторое упорядочение альтернатив вдоль так называемой оси (понятия расстояние между альтернативами не вводится). Каждый агент имеет идеальную альтернативу на этой оси. Если две альтернативы расположены по одну сторону от идеальной альтернативы, то та альтернатива, которая находится ближе к идеальной точке, предпочтается альтернативе, находящейся дальше. Однопиковые предпочтения имеют массу приложений в экономике, психологии, политических науках. Причиной этого являются теоретические свойства однопиковых предпочтений. Во-первых, однопиковые предпочтения хорошо интерпретируемы и позволяют делать удобную визуализацию. В политической науке распространена шкала правых-левых кандидатов, в экономике альтернативой может быть уровень налога или любая другая скалярная величина, в маркетинге агентов различают по лояльности к тому или иному бренду таким образом, что они выстраиваются в линию. Во-вторых, теорема о медианном избирателе [38], гласящая о том, что политическая позиция кандидатов при свободной конкуренции будет стремиться к позиции медианного избирателя, дает дополнительное

обоснование правилу простого большинства. В-третьих, многие вычислительные проблемы теории коллективного выбора значительно упрощаются при однопиковых предпочтениях [39]. В-четвертых, однопиковые предпочтения имеют аксиоматическое обоснование [40]. Домен однопиковых предпочтений для каждой альтернативы имеет хотя бы один линейный порядок с данной альтернативой на первом месте. Все линейные порядки в этом домене связаны друг с другом, то есть один линейный порядок можно получить из другого с помощью последовательного обращения пар соседних альтернатив, при этом все промежуточные линейные порядки тоже принадлежат тому же домену. Кроме того, в этом домене существует пара линейных порядков, такая что каждый из этой пары имеет обратный порядок альтернатив по отношению ко второму. Таким образом, домен однопиковых предпочтений состоит из достаточно близких линейных порядков, но в то же время достаточно разнообразен, так как содержит пару противоположных линейных порядков.

Однолунковые предпочтений являются полной противоположностью однопиковых предпочтений. Все альтернативы также находятся на оси, только каждый агент имеет свою наихудшую альтернативу на этой оси и чем дальше от этой альтернативы, тем лучше. В экономике данные предпочтения встречаются при моделировании предпочтений относительно общественного антиблага (например, предпочтения местных жителей относительно расположения на побережье загрязняющего порта, предприятия и т.д.). Далее в статье однолунковые предпочтения исследоваться не будут. Анализ, проведенный для однопиковых предпочтений можно применить к однолунковым.

В теории коллективного выбора задачи нахождения вероятности того или иного события очень популярны и составляют отдельное направление исследований. Для этого есть множество причин [41]. Теория коллективного выбора полна парадоксов и теорем о невозможности, но в силу отсутствия достаточного количества эмпирических данных мы не можем оценить их частоту на практике. Теоретические оценки позволяют узнать возможности наступления



того или иного события. В частности, даже при достаточно грубых предпосылках очень малая вероятность может сигнализировать о практической невозможности наблюдать некоторый парадокс, свойство и т.д. Несмотря на то, что абсолютные величины теоретических вероятностей сложно интерпретировать, мы можем сравнивать вероятности между собой и делать выводы об их относительной частоте. Имея теоретические формулы для вероятностей, мы можем провести анализ методом сравнительной статистики (к каким изменениям приводит изменение того или иного параметра, например, числа агентов или числа альтернатив).

Наиболее популярными объектами для нахождения вероятностей являются нахождение вероятности парадокса Кондорсе, парадокса двухэтапного голосования (парадокс референдума), манипулируемость правил голосования, и т.д. [42].

Существует множество способов задания функции распределения вероятности. Элементарным событием может быть появление линейного порядка из множества возможных. Определив независимо предпочтения каждого агента, мы получим профиль предпочтений. Такой подход называется основной равновероятной моделью (Impartial Culture, далее IC) Распределение вероятностей может быть построено на ситуациях голосования. Под ситуацией мы можем понимать разбиение числа агентов на группы с одинаковыми предпочтениями, при этом не имеет значение какие предпочтения у конкретного агента (Impartial Anonymous Culture, далее IAC) или дополнительно не обращать внимания на названия альтернатив (Impartial Anonymous Neutral Culture, далее IANC). Равновероятность и независимость являются очень строгими предпосылками, мало общего имеющими с реальными коллективами. Некоторые подходы моделируют скоррелированные предпочтения. Двойственная модель, представленная в статье, моделирует поляризованные предпочтения.

Нахождение вероятности структурированных предпочтений является актуальной темой. Первые результаты для однопиковых предпочтений получены Лакнер и Лакнером [43], для разделимых предпочтений получены Карповым [44]. В данном подразделе приведены результаты, которые углубляют и расширяют результаты, полученные в вышеуказанных работах.

### 1.3.2 Основные понятия

Пусть множество  $X = \{1, \dots, m\}$  будет множеством альтернатив, а множество  $\mathcal{N} = \{1, \dots, n\}$  - множеством агентов. Каждый агент  $i \in \mathcal{N}$  имеет линейный порядок предпочтений  $P_i$  на множестве  $X$ . Первый элемент в этом порядке является наилучшим, последний – наихудшим. Пусть  $\mathcal{L}(X)$  будет множеством всех возможных линейных порядков на множестве  $X$ . Кортеж из  $n$  линейных порядков является профилем предпочтений:  $\mathcal{P} = (P_1, \dots, P_n) \in \mathcal{L}(X)^n$ . Предпочтения будем обозначать как  $a, b, c$ , а множества предпочтений как  $A, B, C$ .

Доменом предпочтений называется множество линейных порядков. Будем говорить, что профиль предпочтений принадлежит домену, если все предпочтения всех агентов принадлежат домену.

Под доменом однопиковых предпочтений будем понимать максимальный домен однопиковых предпочтений. При добавлении к такому домену любого дополнительного линейного порядка домен теряет свойство однопиковости. Домены однопиковых предпочтений имеют  $2^{m-1}$  предпочтений.

Каждый домен однопиковых предпочтений задается осью. Например, ось 123 задает домен  $\{123, 213, 231, 321\}$ . Если у агента альтернатива 1 является наилучшей, то в силу однопиковости он может иметь только линейный порядок 123. Агент с наилучшей альтернативой 2 может далее предпочитать как 1 так и 3. Ось с прямым порядком альтернатив эквивалентна оси с обратным порядком альтернатив. Альтернативы между идеальной для данного агента и альтернативой на конце оси четко упорядочены по предпочтениям: чем ближе к идеальной, тем лучше. Альтернативы с разных сторон от идеальной могут быть упорядочены любым способом.

Таким образом, существует  $m!/2$  различных осей и доменов однопиковых предпочтений соответственно. Оси будем обозначать как  $a, b, c$ .

Баллестер и Хэрингер [45] определили однопиковые предпочтения через запрещенные комбинации предпочтений. В частности, для однопиковых предпочтений требуется, чтобы для любых трёх альтернатив множество наихудших альтернатив может составлять не более двух альтернатив.

### 1.3.3 Вероятностные модели

#### 1.3.3.1 Основная равновероятностная модель (IC)

Существует  $(m!)^n$  разных профилей предпочтений. Все они равновероятны. Для расчета вероятности однопиковых (разделимых) предпочтений необходимо найти число профилей предпочтений с заданными свойствами. Пусть  $\#_{IC}(SP, n, m)$ ,  $\#_{IC}(GS, n, m)$  – количество профилей однопиковых и разделимых предпочтений соответственно с  $n$  агентами и  $m$  альтернативами.

В формулировке приведенной ниже Теоремы 1.12 фигурирует значения  $\#_{IC}(SP, n, m)$  для малого числа альтернатив. Формулы для  $m = 3$  и  $m = 4$  были уже известны, а для  $m = 5$  – получена впервые. Важность нахождения данной формулы обусловлено значительным количеством симуляционных исследований в теории коллективного выбора, которые обычно ограничиваются случаем  $m = 5$  (например [21]).

Теорема 1.12 [46]. Число однопиковых профилей предпочтений равно

$$\#_{IC}(SP, n, 3) = 3 \cdot 4^n - 3 \cdot 2^n;$$

$$\#_{IC}(SP, n, 4) = 12 \cdot 8^n - 18 \cdot 4^n;$$

$$\#_{IC}(SP, n, 5) = 60 \cdot 16^n - 90 \cdot 8^n - 60 \cdot 6^n + 60 \cdot 4^n.$$

Число профилей предпочтений является линейной функцией относительно нескольких показательных функций, где основанием является число линейных порядков в перемечении доменов. Для  $m = 5$ , однопиковые домены могут иметь 2 линейных порядка в пересечении, но в силу формулы включения исключения коэффициент перед этим членом равен нулю.

### 1.3.3.2 Анонимная модель (IAC)

Согласно анонимной модели профили эквивалентны, если один можно получить из другого с помощью перестановки (переименования) агентов. Все данные профили предпочтений объединены в ситуацию голосования. Ситуации голосования распределены равновероятно. Впервые модель была предложена в работе [47] как развитие работы [48].

Число ситуаций голосования равно числу мультимножеств из  $n$  элементов на множестве  $m!$  линейных порядков:  $\binom{m!+n-1}{n}$ . Число ситуаций голосования с  $k$  линейными порядками равно  $\binom{k+n-1}{n}$ . Следуя той же логике включения исключения как при выводе утверждений 1 и 2 мы можем получить число однопиковых ситуаций голосования, заменяя  $k^n$  на  $\binom{k+n-1}{n}$ .

### 1.3.3.3 Равномерная модель (Uniform culture)

Герлейн [49] предложил равномерную модель для  $m = 3$ . В этой модели 6 событий появления разных линейных порядков распределены равномерно на соответствующем симплексе. Имея 3 домена с 4 линейными порядками, каждая пара из которых имеет пересечение в два линейных порядка, получим вероятность однопиковых профилей предпочтений

$$\frac{3 \int_{S_5} ((p_1 + p_2 + p_3 + p_4)^n - (p_1 + p_2)^n) d\mathbf{p}}{m(S_5)},$$

где  $S_k = \left\{ \mathbf{p} = (p_1, \dots, p_k) : p_i \in [0, 1], \sum_{i=1}^k p_i \leq 1 \right\}$  –  $k$ -мерный симплекс,  $m(S_k)$  – его мера, что равно

$$Pr_{UC}(SP, n, 3) = Pr_{UC}(GS, n, 3) = \frac{60n}{(n+2)(n+3)(n+4)}.$$

Данная вероятность равна соответствующей вероятности для анонимной модели.

#### 1.3.3.4 Двойственная модель

Для моделирования поляризованных предпочтений (подробнее о поляризованных предпочтениях в [50]) мы вводим зависимость между вероятностями линейных порядков:  $p_1 = \Pr(123) = \Pr(321)$ ,  $p_2 = \Pr(132) = \Pr(231)$ ,  $p_1 = \Pr(213) = \Pr(312)$ .

Так как объемы всех доменов и их пересечений равны, то вероятность однопиковых предпочтений равна

$$Pr_2(SP, n, 3) \frac{\int_0^{0.5} \int_0^{0.5-p_1} 3[0.5+p_1]^n - 3(p_1+p_2)^n dp_2 dp_1}{\int_0^{0.5} \int_0^{0.5-p_1} 1 dp_2 dp_1} = \frac{24 - (12n+24)(0.5)^n}{(n+1)(n+2)}.$$

Результаты данного подраздела в полном объеме отражены в [46]. Полученные вероятности оказались малы, что соответствует эмпирическим данным. Современным решением практической применимости однопиковых предпочтений является изучение около однопиковых предпочтений ([51], [52]). Нахождение вероятности около однопиковых профилей предпочтений является естественным развитием данной работы.

### 1.4 Исследование турнирных способов выбора оптимальных вариантов

Основной задачей теории коллективного выбора является описание способов определения альтернатив, которые должны быть выбраны из числа имеющихся в наличии вариантов на основании мнения о них индивидуальных участников процесса принятия коллективных решений.

Математически выбор моделируется функцией выбора. Были рассмотрены четыре схожие функции выбора, зависящие от коллективных предпочтений, и основанные на понятии устойчивости множества: объединение минимальных доминирующих множеств  $D$ , объединение минимальных слабоустойчивых множеств  $WS$ , объединения минимальных  $P$ - и  $R$ -внешнеустойчивых множеств  $ES$  и  $RES$ .

Ключевая проблема, рассматривавшаяся в данной части исследования, — это проблема пустоты (то есть невозможности) оптимального выбора в случае

бесконечного множества вариантов, то есть проблема отсутствия альтернатив, удовлетворяющих критериям оптимальности, воплощённым в этих решениях.

#### 1.4.1 Описание модели

Рассматривается предъявление – множество  $X$  альтернатив, доступных для выбора. Общество принимает коллективные решения, которые сводятся к выбору альтернатив из  $X$ . В исследовании предполагается, что на основании профиля индивидуальных предпочтений членов общества можно определить отношение  $P$ ,  $P \subseteq X \times X$ , моделирующее строгие коллективные предпочтения. Коллективный выбор моделирует функция  $SC(P): 2^{X \times X} \rightarrow 2^X$ , аргументом которой являются коллективные предпочтения, а значениями – подмножества предъявления, соответствующие выбранным альтернативам. Бинарное отношение  $P$  называют турниром, а коллективный выбор на его основании эквивалентен выбору победителя в турнире.

Если функция выбора  $SC(P)$  удовлетворяет аксиомам нейтральности (независимости выбора от имён альтернатив) и непустоты выбора, а также соответствует принципу Кондорсе, то её называют *турнирным решением* [53].

"Наивное" определение оптимального коллективного выбора по аналогии с оптимальным индивидуальным выбором, как выбора  $P$ -недоминируемой альтернативы в общем случае невозможно, так как таких альтернатив может не быть. Этот результат известен как «парадокс Кондорсе» [54]. Из-за него приходится вводить иные, нетривиальные принципы выбора наилучших вариантов. Исследуемые решения основаны на принципе выбора альтернатив, принадлежащих множеству с каким-либо «хорошим» свойством.

Подмножество  $Y$  предъявления  $X$  называется доминирующим, если  $\forall x \in X, \exists y \in Y: yPx$  [55],  $P$ -внешнеустойчивым, если  $\forall x \in X \setminus Y, \exists y \in Y: yPx$  [56],  $R$ -внешнеустойчивым, если  $\forall x \in X \setminus Y, \exists y \in Y: yRx$  [57], слабоустойчивым, если  $\forall x \in X \setminus Y, \forall z \in Y, xPz \Rightarrow (\exists y \in Y: yPx)$  [19].

Множество называется минимальным относительно некоторого свойства, если ни одно из его подмножеств, кроме него самого, этим свойством не

обладает. Наилучшей считается альтернатива, принадлежащая хотя бы одному минимальному множеству, поэтому решениями являются объединение минимальных доминирующих множеств  $D$ ; объединение минимальных  $P$ -внешнеустойчивых множеств  $ES$ , объединение минимальных  $R$ -внешнеустойчивых множеств  $RES$  и объединение минимальных слабо-устойчивых множеств  $WS$ .

Идея выбора минимального  $P$ -внешнеустойчивого множества впервые встречается в [58]. Однако авторы никак её не развивают и не используют. Независимо от них и в другом контексте  $MPES$  было определено Субочевым [59] по аналогии с минимальным слабоустойчивым множеством, предложенным Алескеревым и Курбановым [19]. Минимальное  $R$ -внешнеустойчивое множество было введено Алескеревым и Субочевым [57], [60] по аналогии с  $MPES$ . Минимальное доминирующее множество было введено в предшествующей части настоящего исследования [61].

#### 1.4.2 Характеристические теоремы

На более ранних этапах исследования были получены характеристические теоремы, [57] - [61], связывающие три рассматриваемых решения с другим широко известным турнирным решением - непокрытым множеством [62], [63].

Пусть  $Q$  – бинарное отношение на  $X$ ,  $Q \subseteq X \times X$ . Множество альтернатив, которые  $Q$ -доминируют альтернативу  $x$ , называется её *верхним срезом*  $Q(x) = \{y \in X \mid yQx\}$ . Множество альтернатив, которые  $Q$ -доминируемы альтернативой  $x$ , называется её *нижним срезом*  $Q^{-1}(x) = \{y \in X \mid xQy\}$ .

*Отношение покрытия*  $C(P)$  и по Миллеру, и по Фишберну есть усиленная версия строгих коллективных предпочтений  $P$ :  $C(P) \subseteq P$ .

Альтернатива  $x$  *покрывает* альтернативу  $y$  по Миллеру, если  $x$  строго лучше  $y$ , и любая альтернатива предъявления, которая строго хуже  $y$ , также строго хуже  $x$ :  $x C_M(P) y \Leftrightarrow x P y \wedge P^{-1}(y) \subset P^{-1}(x)$ .

Альтернатива  $x$  покрывает альтернативу  $y$  по Фишберну, если  $x$  строго лучше  $y$ , и любая альтернатива предъявления, которая строго лучше  $x$ , также строго лучше  $y$ :  $xC_F(P)y \Leftrightarrow xPy \wedge P(x) \subset P(y)$ .

Альтернатива  $x$  нестрого покрывает альтернативу  $y$  по Миллеру  $xC_{WM}(P)y \Leftrightarrow P^{-1}(y) \subset P^{-1}(x)$ .

Альтернатива  $x$  нестрого покрывает альтернативу  $y$  по Фишберну  $xC_F(P|_A)y \Leftrightarrow P(x) \subset P(y)$ .

Множество альтернатив не покрытых в  $X$  никакими альтернативами называется *непокрытым множеством* предъявления  $X$ . Множество непокрытых в  $X$  альтернатив, которые также не покрыты в  $X$  никакими альтернативами нестрого, называется *внутренним непокрытым множеством* предъявления  $X$ . Непокрытые множества по Миллеру и Фишберну, а также их внутренние версии обозначаются  $UC_M$  и  $UC_F$ ,  $UC_{IM}$  и  $UC_{IF}$ , соответственно.

Теорема 1.13. Пусть  $X$  конечно. Альтернатива  $x$  принадлежит объединению минимальных  $P$ -внешнеустойчивых множеств  $ES$  в том и только в том случае, если она строго лучше некоторой альтернативы из непокрытого множества Фишберна или сама принадлежит к этому множеству,

$$x \in ES \Leftrightarrow \exists y: y \in P^{-1}(x) \cap UC_F \vee x \in UC_F.$$

Теорема 1.14. Пусть  $X$  конечно. Альтернатива  $x$  принадлежит объединению минимальных  $R$ -внешнеустойчивых множеств  $RES$  в том и только в том случае, если она не хуже некоторой альтернативы из внутреннего непокрытого множества Миллера,

$$x \in RES \Leftrightarrow \exists y: y \in R^{-1}(x) \cap UC_{IM}.$$

Теорема 1.15. Пусть  $X$  конечно. Альтернатива  $x$  принадлежит объединению минимальных доминирующих множеств  $D$  в том и только в том случае, если она строго лучше некоторой альтернативы из внутреннего непокрытого множества Фишберна,

$$x \in D \Leftrightarrow \exists y: y \in P^{-1}(x) \cap UC_{IF}.$$



Эти теоремы дают удобный способ вычислять вышеназванные решения. Его описание можно найти в [60].

### 1.4.3 Проблема выбора из бесконечного множества

Ключевым предположением характеристических теорем является конечность предъявления. Во-первых, конечность  $X$  обеспечивает непустоту всех версий непокрытого множества. Во-вторых, она гарантирует, что в любом множестве, наделённом некоторыми свойствами, есть подмножество, минимальное в отношении данного свойства. Как показывает следующий пример, эти утверждения могут быть неверными, когда число альтернатив в предъявлении бесконечно. Пусть  $X$  – это (бесконечная) последовательность  $\{x_n\}$ ,  $n=1, 2, 3, \dots$ . Пусть  $x_n P x_m \Leftrightarrow n > m$ . Отношение  $P$  – турнирное и линейное решение. Очевидно, что  $x_n C x_m \Leftrightarrow n > m$ . Следовательно,  $UC = \emptyset$ . Любое бесконечное подмножество  $X$  доминирующее и, следовательно, внешнеустойчивое. Но любое конечное подмножество  $X$  не внешнеустойчиво и, следовательно, не является доминирующим. Так как в любом бесконечном множестве есть несовпадающее с ним бесконечное подмножество, в  $X$  нет ни одного внешнеустойчивого или доминирующего множества, которые были бы минимальными.

Если в предъявлении бесконечно много альтернатив, утверждения Теорем 1.13-1.15 могут быть неверны. Например, пусть  $X$  – это объединение трёх (бесконечных) последовательностей  $\{x_n\}$ ,  $\{y_n\}$  и  $\{z_n\}$ ,  $n=1, 2, 3, \dots$ . Пусть  $x_n P y_m \Leftrightarrow n > m$ ;  $y_n P y_m \Leftrightarrow n > m$ ;  $z_n P z_m \Leftrightarrow n > m$ ;  $\forall n, \forall m, x_n P y_m \wedge y_n P z_m \wedge z_n P x_m$ . Тогда  $UC = \emptyset$  и любая тройка  $\{x_n, y_m, z_k\}$  является минимальным доминирующим и минимальным внешнеустойчивым множеством; следовательно,  $D = ES = X$ , что противоречит основному утверждению всех трёх теорем.

Достаточное условие непустоты и  $C_{WF}$ -внешней устойчивости внутреннего непокрытого множества по Фишберну установлено в [64],[65].

Теорема 1.16 [64],[65]. Если на  $X$  можно ввести топологию так, чтобы  $X$  было в ней компактным, а множества  $P^{-1}(x)$  – открытыми для всех  $x \in X$ , то  $UC_{IF} \neq \emptyset$  и  $\forall y \notin UC_{IF}, \exists x \in UC_{IF}: x C_{WF} y$ .

Оказывается, что этого условия также достаточно для существования непустых минимальных доминирующих и минимальных внешнеустойчивых множеств.

Лемма 1.1. Если на  $X$  можно ввести топологию так, чтобы  $X$  было в ней компактным, а множества  $P^{-1}(x)$  – открытыми для всех  $x \in X$ , и если множество максимальных элементов отношения  $P$  пусто, то в любом доминирующем множестве есть непустое, конечное, доминирующее подмножество.

Следствие 1.1. Если выполняются условия Леммы 1.1, все минимальные доминирующие множества будут конечными, и в каждом доминирующем множестве будет конечное минимальное доминирующее подмножество, а, следовательно, и минимальное  $P$ -внешнеустойчивое подмножество, минимальное  $R$ -внешнеустойчивое подмножество и минимальное слабоустойчивое подмножество.

Теорема 1.17. Если на  $X$  можно ввести топологию так, то  $X$  будет в ней компактным, а множества  $P^{-1}(x)$  – открытыми для всех  $x \in X$ , тогда  $D \neq \emptyset$ ,  $RES \neq \emptyset$  и  $WS \neq \emptyset$ . Если к тому же множество максимальных элементов отношения  $P$  пусто или внешнеустойчиво, тогда и  $ES \neq \emptyset$ .

Также появляется возможность обобщить утверждение Теоремы 1.15.

Теорема 1.18. Если на  $X$  можно ввести топологию так, чтобы  $X$  было в ней компактным, а множества  $P^{-1}(x)$  – открытыми для всех  $x \in X$ , то  $D \neq \emptyset$ ,  $RES \neq \emptyset$  и  $WS \neq \emptyset$ . Если к тому же множество максимальных элементов отношения  $P$  пусто, то  $x \in D \Leftrightarrow \exists y \in UC_{\text{IF}}: xPy$ .

Нахождение условий, при которых утверждения Теорем 1.14 и 1.15 остаются справедливыми и в случае выбора из бесконечного множества, является интересной открытой проблемой, которая будет рассмотрена на следующем этапе исследования.

## 1.5 Анализа предпочтений студентов относительно форматов онлайн-обучения

### 1.5.1 Введение

Онлайн образование стремительно набирает свою популярность. Начиная с 2010 годов, в Интернете впервые стали появляться платформы с массовыми открытыми онлайн-курсами (сокращенно MOOC), например, Coursera, edX, Национальная платформа открытого образования и др. По данным Class Central<sup>2</sup>, одного из самых известных и авторитетных MOOC-агрегаторов, число людей, зарегистрированных как минимум на один курс, составляет более 110 миллионов, а количество курсов увеличилось до 13500 (рисунок 1.11).

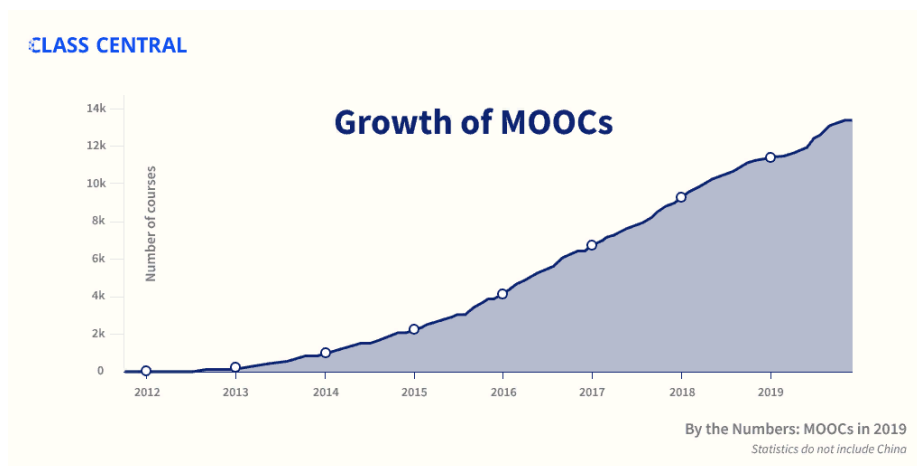


Рисунок 1.11 – Рост онлайн-образования

Все больше онлайн-обучение внедряется в университетах. В частности, в последние годы учебные планы студентов НИУ ВШЭ обязательно включается онлайн-курс (в формате массового открытого онлайн-курса или дисциплин в формате blended learning, когда студенты изучают часть предмета дистанционно). Так, в 2018–2019 учебном году почти 17 тысяч студентов НИУ ВШЭ изучали 443 массовых открытых онлайн-курса. Ректор НИУ ВШЭ

---

<sup>2</sup> <https://www.classcentral.com/report/MOOCs-stats-and-trends-2016/>

Я.И.Кузьминов<sup>3</sup> в интервью 2018 года сказал: «Мы отменим лекционные занятия в классической форме. Думаю, что через пять лет мы точно заменим абсолютно все лекции онлайн-курсами. То есть каждый профессор "Вышки" запишет и будет поддерживать онлайн-курс по своему предмету».

Пандемия коронавируса и введение карантина во многих странах мира резко ускорила эту тенденцию. Практически все вузы и школы стали использовать элементы онлайн-обучения. Такой вынужденный массовый переход привлек большое внимание к недостаткам онлайн-образования, среди которых наиболее важными являются отсутствие прямого контакта и взаимодействия студента с преподавателем, а также невозможность полностью контролировать самостоятельность выполнения оценочных заданий студентом.

Данная работа была начата еще до пандемии и массового перехода студентов на онлайн-обучение по причине все большего использования элементов онлайн-образования в традиционном формате университетского обучения и заявленной тенденции полного перехода в будущем на подобный формат. Целью этой работы является анализ мнения студентов ВШЭ относительно разных форматов онлайн-обучения для выявления причин и факторов, влияющих на их отношение к онлайн формату с помощью методов латентных классов и дисперсионного анализа.

#### 1.5.2 Основная часть

Для проведения исследования были взяты данные «Мониторинга студенческой жизни» за 2018 и 2019гг. Мониторинг студенческой жизни – это ежегодный опрос в ВШЭ, проводимый Центром внутреннего мониторинга, на основе которого выявляется отношение студентов к качеству образования в вузе, к созданным Вышкой условиям обучения и проживания, к вовлеченности в студенческую жизнь и др. За 2018 год есть данные с ответами 6631 студента,

---

<sup>3</sup> <https://www.kommersant.ru/doc/3758336>

за 2019 год – 5688 ответов студентов. Для нашего исследования потребуется только часть вопросов из этой анкеты, а именно, нам важно составить портрет студентов, проходивших опрос, включающий данные о курсе обучения, месте проживания, гражданстве, наличии работы, материальном положении. Основными вопросами являются, конечно, вопросы об онлайн курсах и предпочтениях относительно онлайн формата обучения. Выбранные вопросы анкеты приведены в Таблице А.1 Приложения А.

Краткий ответ по отношению студентов к онлайн-образованию по опросу 2018 года и 2019 года можно получить из нижеследующих Таблиц 1.3 и 1.4. В первой приведено распределение ответов по вопросу анкеты 2018 года: «Если Вы в прошлом или текущем учебном году включали в свой учебный план дисциплины онлайн или с элементами онлайн, то какой формат учебного курса предпочтителен для Вас: онлайн-курс или курс, предполагающий аудиторные занятия с преподавателем?». Во второй – распределение ответов по вопросу анкету 2019 года: «При прочих равных условиях какой формат учебного курса Вы бы выбрали?».

Таблица 1.3 – Результаты опроса 2018 года

	Вариант ответа	Процент студентов
1	Определенно предпочитаю онлайн-курсы	3,06%
2	Скорее предпочитаю онлайн-курсы	3,62%
3	Одинаково отношусь к обоим форматам	7,86%
4	Предпочитаю курсы, сочетающие в себе онлайн занятия с аудиторными занятиями с преподавателем	14,12%
5	Скорее предпочитаю курсы, предполагающие аудиторные занятия с преподавателем	10,99%
6	Определенно предпочитаю курсы, предполагающие аудиторные занятия с преподавателем	12,88%
7	Не отвечали на этот вопрос, так как не проходили онлайн курсов или смешанных курсов	46,18%

Таблица 1.4 – Результаты опроса 2019 года

	Вариант ответа	Процент студентов
1	Онлайн-курсы	10,7%
2	Дисциплины со смешанным обучением	44,99%
3	Оффлайн-курсы	44,3%
4	Мне всё равно	0%

Далее был произведен кластерный анализ методом латентных классов, чтобы выделить однородные группы студентов и понять их мотивы поддержки или, наоборот, нежелания поддерживать переход на онлайн-обучение. Этот метод не распределяет объекты однозначно по конкретным кластерам, а для каждого кластера рассчитывает вероятность попадания в него объектов. В этом случае, отнеся объекты к кластерам с наибольшей вероятностью попадания, можно получить однозначное разделение на кластеры, как и во всех описанных выше методах. Однако преимущество данного способа кластеризации в том, что наглядно видно, насколько вероятно было попадание объекта в кластер и насколько однозначно полученное распределение.

Вероятности принадлежности к кластеру вычисляются по формуле:

$$P(Y = y) = \sum_{x=1}^C P(X = x) \prod_{l=1}^L P(Y_l = y_l | X = x),$$

где  $L$  – количество переменных,  $y_l$  – ответы на вопрос  $Y_l$ ,  $C$  – количество латентных классов,  $x$  – латентный класс,  $X$  – латентная переменная.

После подсчета вероятностей по описанной формуле оцениваются параметры полученной модели и в результате получается разделение всех объектов на конкретное число кластеров.

Сначала опишем результаты кластеризации по данным опроса 2018 года. Результаты приведены в Таблице 1.5, в которой указаны основные характеристики полученных кластеров.

Таблица 1.5 – Описание латентных классов, вопросы про онлайн формат

Кластер	Брали ли онлайн курсы	За онлайн, или смешанный, или аудиторный формат	Отношение к высказываниям против онлайн формата	Отношение к переводу лекций в онлайн
1	никто не брал	нет ответов	нет ответов	за
2	все брали (онлайн и смешанные)	смешанный, аудиторный	больше не согласны	за
3	никто не брал	нет ответов	нет ответов	против
4	все брали (онлайн и смешанные)	смешанный	больше не согласны	против
5	никто не брал	нет ответов	нет ответов	за
6	большинство брали (онлайн)	смешанный, аудиторный	одинаковое распределение между ответами	скорее против

Студенты, попавшие в кластеры 1, 3 и 5, никогда не включали в свою учебную программу ни полностью онлайн, ни смешанные курсы, поэтому они не отвечали на вопросы о предпочтительном формате обучения и не высказывали свое отношение к ним. Несмотря на это, они высказали свое мнение относительно замены лекций онлайн курсами, и мнения разделились: студенты из кластеров 1,2 и 5 выступают «за» перевод традиционных лекций в формат онлайн, а кластеры 3 и 4 оказались «против», причем по данному вопросу кластеры разделились таким образом, что в каждом кластере все студенты либо принимают позицию противников перевода, либо сторонников, только в кластере 6 присутствуют и сторонники, и противники, но все же число противников перевешивает. В отличие от кластеров 1, 2, 5, все люди из кластеров 2 и 4 имели опыт обучения по онлайн курсам и не согласны с высказываниями против онлайн курсов, они отмечают только, что список онлайн курсов достаточно короткий. Значит, студенты данных кластеров удовлетворены курсами, при этом в кластере 2 все выступают «за» замену лекций онлайн курсами, а в кластере 4 – «против».

Как можно судить по полученной кластеризации, в 6 кластер вошли все те, кто не подошел по ответам к остальным кластерам. Там примерно половина участников имеют опыт обучения по онлайн курсам, разошлись во мнениях относительно высказываний против онлайн обучения и скорее выступают против перевода лекций на онлайн курсы.

Таким образом, и среди магистров, и среди бакалавров есть как сторонники перевода на онлайн курсы, так и противники, а также не наблюдается прямой зависимости между переводом и наличием опыта обучения по онлайн курсам.

Теперь для каждого кластера проверим, какие вопросы и ответы оказывали наибольшее влияние на отношение студентов к переводу лекций в онлайн формат. Для этого в каждом сформированном кластере мы проверяем значимость вопросов, чтобы выделить из них наиболее значимые, сформировать характеристику кластера и выявить возможные причины предпочтений студентов. Данная проверка зависимостей будет осуществляться с помощью дисперсионного анализа ANOVA, который позволяет исследовать значимости отличий в средних значениях групп с помощью сравнения дисперсий. Данный метод реализован в SPSS, поэтому все вычисления производились в этой программе. Результаты представлены в Таблице А.2 Приложения А.

Проанализировав данную таблицу, можно охарактеризовать каждый кластер. Можно предположить, что в кластер 1 попали очень преданные Вышке студенты, которые высоко ценят качество обучения, поэтому они выступают «за» любые нововведения вуза, а в данном случае перевод лекций в онлайн формат. В кластере 2 студенты, любители онлайн курсов: возможно, им очень удобен такой формат обучения в связи с какой-то дополнительной занятостью, а может им просто комфортно воспринимать материал дома с ноутбуком, поэтому они тоже поддерживают идею перевода. Люди из кластера 3, скорее всего очень дисциплинированные, которые любят, чтобы все было четко по графику и внимательно слушают преподавателя на парах, не отвлекаясь на посторонние вещи, поэтому идея смены формата не подходит под их взгляды и привычный образ жизни. В кластер 4 попали социальные люди, которые любят общаться и взаимодействовать с людьми, в том числе им так проще учиться, в связи с этим данная группа людей также против перехода на онлайн формат. Людей из кластера 5 можно назвать студентами, учащимися



только ради оценок, так как самым значимым параметром стал вопрос о переводе оценок. Проблем с переводом оценок нет, поэтому студенты также выступают «за» онлайн формат. Как упоминалось ранее, кластер 6 можно описать как смешанный и разнообразный.

Теперь опишем результаты кластеризации по данным опроса 2019 года. Результаты приведены в Таблице А.3 Приложения А. Здесь была выбрана модель с восемью кластерами по критерию ВИС (байесовскому информационному критерию).

Студенты, попавшие в кластеры 1 и 8, никогда не включали в свою учебную программу ни полностью онлайн, ни смешанные курсы, поэтому они не отвечали на вопросы о предпочтительном формате обучения и не высказывали свое отношение к ним. Все студенты из кластера 2 имели опыт обучения по онлайн курсам и не брали курсы в смешанных форматах, при этом они не сталкивались с трудностями во время прохождения онлайн-курсов, однако подавляющее большинство (73%) предпочитают оффлайн или смешанные курсы, за онлайн проголосовало всего 14%. В кластере 3 все студенты брали смешанные курсы (с онлайн-лекциями и оффлайн-семинарами), а также большинство брали и онлайн-курсы; при этом они также не отметили трудностей и большинством в 84% за оффлайн или смешанный формат, всего 6% за онлайн-курсы. Кластеры 4 и 5 похожи – в них все или почти все студенты имеют опыт прохождения онлайн-курсов и немного знакомы со смешанным форматом (в виде онлайн-лекций без оффлайн-семинаров, очно только экзамен), при этом все они отмечают недостатки онлайн-курсов и в большинстве поддерживают либо оффлайн, либо оффлайн и смешанный формат. Шестой кластер состоит только из студентов, бравших смешанные курсы второго типа (онлайн-лекции без оффлайн-семинаров, очно только экзамен), треть из них при этом отмечают такие недостатки как «узкий список онлайн-дисциплин» и «мне было непонятно, у кого можно проконсультироваться» и также почти все за оффлайн и смешанный формат. В седьмом кластере находятся студенты, все бравшие смешанные курсы с онлайн-лекциями и оффлайн-семинарами, большинство

из которых при этом также брали и онлайн-курсы, но при отсутствии замеченных ими трудностей также предпочитают смешанный формат. В 2019 году в анкете не задавался вопрос об отношении к переводу лекций в онлайн.

Укажем также некоторые особенности кластеров. Так, в кластеры 1, 3 и 5 больше всего первокурсников бакалавриата (54%, 58% и 41%), а магистры распределились в основном в кластерах 7 и 8. В кластерах 7 и 8 больше всего (32% и 40%) работающих на полный рабочий день (причем вне Вышки) студентов. Во всех кластерах, кроме 4, очень высока доля ответов о выборе такой же образовательной программы в НИУ ВШЭ как альтернативного решения о поступлении, и только в 4 кластере лишь половина поддерживает такое решение, четверть студентов выбрала бы или другую программу или другой вуз.

Теперь для каждого кластера проверим, какие вопросы и ответы оказывали наибольшее влияние на отношение студентов к переводу лекций в онлайн формат. Для этого в каждом сформированном кластере мы проверяем значимость вопросов, чтобы выделить из них наиболее значимые, сформировать характеристику кластера и выявить возможные причины предпочтений студентов с помощью дисперсионного анализа ANOVA. Поскольку в опросе 2019 года не было обязательного вопроса о переходе на онлайн формат, то будем оценивать предпочтения студентов по вопросу о предпочтительной форме онлайн-курсов. При этом, однако, такой анализ будет невозможен для кластеров 1 и 8, в которых находятся студенты, не изучавшие онлайн-курсы и не отвечавшие на этот вопрос. Результаты представлены в Таблице А.4 Приложения А.

### 1.5.3 Некоторые выводы

Метод анализа латентных классов произвел достаточно четкое разделение на кластеры по ответам студентов, благодаря чему стало возможно сделать некоторые выводы. Прежде всего, по результатам кластеризации стало понятно, что студенты имеют схожие предпочтения по поводу обучения в Вышке – по многим вопросам, указанным выше, кластеры между собой ни-

как не отличались. Но с помощью анализа латентных классов и дисперсионного анализа удалось разбить данные на четко разделяющиеся кластеры, отличающиеся между собой отношением к переводу традиционных лекций в онлайн формат, и причинами такого отношения.

К сожалению, по данным опроса 2019 года не получается выделить столь же однородные и содержательно трактуемые кластеры, как в опросе 2018 года. Однако видно, что на склонность к выбору онлайн-курсов и курсов смешанного формата может оказать влияние недовольство вузом или образовательной программой, психологическое ощущение комфорта от процесса обучения и различные требования университета по формированию образовательного трека.

Результаты проведенных исследований нашли отражение в [66].

## **2 Исследование моделей сетевого анализа данных**

В этом разделе отчета приведены результаты исследований моделей сетевого анализа данных в следующих направлениях:

- разработка обобщенного индекса LRIC, учитывающего влияние в сетевых структурах;
- разработка и применение индексов центральности, учитывающих индивидуальные и групповые влияния;
- оценивание схожести слоёв сети торговли продовольствием;
- исследование многосторонних торговых потоков в гравитационной модели торговли;
- оценивание центральностей многослойной сети методами теории функций доверия.

### **2.1 Введение**

В последнее время большое внимание уделяется анализу различных сообществ и сложных сетей, в частности исследованию их структуры и определению ключевых элементов. На данный момент существует большое количество индексов центральности, с помощью которых можно оценить степень влияния каждого элемента сети. Чем выше центральность вершины, тем более влиятельной она является в сети. Тем не менее, поскольку каждый индекс центральности имеет свою интерпретацию того, какой элемент необходимо считать наиболее важным, данные индексы принимают во внимание различную информацию о характеристиках вершин и структуре их взаимосвязей.

Например, степенная центральность (degree centrality) основана на подсчете общего числа соседей каждой вершины [67]. В зависимости от типа сети данный индекс позволяет учитывать направление взаимосвязей между вершинами (in/out degree), а также степень их интенсивности (weighted degree). Другой группой индексов центральности, основанных на подсчете степеней вершин, являются индексы, основанные на подсчете собственного вектора. Наиболее известными и широко применяемыми являются такие индексы, как  $\alpha$ -centrality [68], Bonacich [69], Katz [70], PageRank [71], Hubs and Authorities

[72]. Основная идея данных индексов заключается в том, что важность вершины зависит от важности ее соседей, которая, в свою очередь, в важности их соседей и т.д.

Следующая группа индексов центральности основана на подсчете кратчайших путей между элементами сети. Наиболее известными представителями данной группы являются индексы центральности по близости (closeness centrality) [73] и посредничеству (betweenness centrality) [74]. Индекс центральности по близости выявляет вершины, которые находятся ближе всего ко всем остальным элементам сети. Обобщение данного индекса было рассмотрено в [75]. Индекс центральности по посредничеству выявляет вершины, которые наиболее часто встречаются на кратчайших путях между любыми другими вершинами. Чаще всего элементы, являющиеся наиболее важными по данному индексу, играют важную роль в связности графа. Различные модификации данного индекса рассмотрены в [76]-[78].

Задача выявления наиболее влиятельных элементов может быть также рассмотрена с точки зрения теоретико-игрового подхода. В таком случае сеть рассматривается как набор взаимодействующих индивидов, которые вносят вклад в некоторое общее дело. Основной проблемой данной задачи является распределение полученного выигрыша между индивидами. В [79] был предложен вектор Майерсона (Myerson value), основанный на индексе Шепли-Шубика [80]. Вектор Майерсона определяет средний вклад каждого участника  $i$  на основе того, насколько увеличивается общий выигрыш участников сети по сравнению с той ситуацией, когда участник  $i$  не принимает свое участие.

К сожалению, большинство из них не позволяют учитывать важные характеристики: во-первых, не учитывались параметры вершин, хотя в фундаментальной работе [81] отмечается этот недостаток. Во-вторых, не принималось во внимание влияние группы вершин на некоторую определенную вершину. Более того, существующие методы не в полной мере оценивают интенсивность дальних взаимодействий и зачастую учитывают незначимые связи между элементами сети. Следовательно, результаты применения классических

индексов центральности не в полной мере отображают реальную картину, так как с их помощью недооценить действительно важных элементов сети.

В работе [82] был предложен новый метод оценки влияния в сетях (SRIC). Отличительной его особенностью является то, что он учитывает такие индивидуальные характеристики вершины, как ее размер (силу) и критическое пороговое значение, при достижении которого другие вершины начинают на нее влиять. Тем не менее, разработанный индекс учитывает исключительно ближние взаимодействия между вершинами, а степень влияния вершин, не имеющих друг с другом прямых связей, оставалось нулевой.

В работе [83], [84] предложен ряд моделей (LRIC), позволяющих учесть существующие недостатки индекса SRIC. Отличительной особенностью данных моделей является то, что они позволяют оценить влияние вершин, не имеющих между собой прямых связей и находящихся на расстоянии  $s$ , значение которого может варьироваться от задачи. Интенсивность дальних взаимодействий может быть оценена как с помощью анализа всех возможных каналов влияния одной вершины на другую, так и с помощью имитационного подхода [85]. Вычислительная сложность данных индексов рассмотрена в [86]. Усовершенствование индексов ближних и дальних взаимодействий было также предложено в работе [87].

Отметим, что в ряде сетевых структур наличие связи между вершинами может приводить к взаимозависимости вершин друг на друга. Например, если рассмотреть торговые отношения между странами, то высокие объемы торговли приводят к зависимости импортера относительно тех продуктов, которые производятся в недостаточном объеме и поэтому закупаются из других стран. Заметим, что в таком случае страна-экспортер может использовать торговлю как элемент давления на страну-импортера. С другой стороны, если страна-импортер неожиданно откажется от покупки товаров, то это также может привести к экономическим потерям страны-экспортера. Таким образом, при некоторых обстоятельствах связь между вершинами приводит к их взаи-

мозависимости. Данная особенность не учитывается существующими индексами центральности, в связи с чем возникает необходимость в разработке новых моделей.

Анализ влияния стран с помощью классических и новых индексов центральности был применён к сети международной миграции, сети общей торговли, сети конфликтов, сети торговли нефтью и газом, а также к сети взаимодействия террористических групп.

Отметим также важность применения сетевого анализа к данным видам сетевых взаимодействий. Например, многие работы в области международной миграции рассматривают миграционные потоки между странами мира независимо друг от друга. Однако международная миграция является одним из аспектов глобализации, и невозможно воспринимать международные миграционные потоки изолированно. Появление данных о двусторонних потоках международной миграции позволило представить процесс международной миграции как систему, в которой все страны взаимосвязаны. Страны мира представлены как вершины сети, а миграционные потоки между ними – как ребра. В работе использованы методы сетевого анализа для оценки влияния стран в международной миграции. Для этого применяются новые индексы центральности, учитывающие как индивидуальное, так и групповое влияние стран в сети.

Что касается сети природных ресурсов, то отметим, что проблема истощения таких невозобновляемых ресурсов как нефть и природный газ всемирно известна и не является решенной на текущий момент. По мере того, как население увеличивается, растет и спрос на сырье, тенденция такова, что с каждым годом мировое потребление нефти, газа и угля увеличивается. В 2018 году доля потребления нефти в мире составила 34% от всех видов топлива, доля потребления природного газа – 24%, на уголь пришлось 27%, менее 15% - на возобновляемые источники и атомную энергию.

В современном мире нефть – это не просто ресурс, а полноценный финансовый инструмент, от которого зависят политика и экономика многих

стран. Мир ориентируется в основном на цены двух марок нефти – Brent Северного моря и West Texas Intermediate (WTI), добываемую в штате Техас, США. Сегодня на цену нефти влияет множество факторов – в первую очередь величины спроса и предложения на мировом рынке, которые проанализированы в работе.

Относительно природного газа, то только за 2018 год его потребление и добыча выросли на более чем 5% – это самый высокий показатель за последние 30 лет. Основной спрос был представлен США и Китаем, в основном это связано с переходом на более экологичные виды топлива. Причина этого кроется в том, что природный газ часто называется «зеленым топливом», так как при его горении образуются только вода и углекислый газ, а во время горения нефти, каменного угля или древесины образуются зола и копоть. Кроме этого, газ как сжиженный, так и сжатый используется как автомобильное топливо, потому что выбросы углекислого газа гораздо меньше в сравнении с остальными энергоресурсами, к тому же он дешевле бензина и соответствует стандарту Евро – 4.

В работе рассмотрены экспорт и импорт трех продуктов: сырая нефть и нефтепродукты, полученные из битуминозных пород (270900 по гармонизированной системе), природный газ сжиженный (271111), природный газ в газообразном состоянии (271121). Также в работе использованы данные о добыче и внутреннем потреблении нефти и газа, предоставленные транснациональной нефтегазовой компанией British Petroleum (BP) [88]. Данные по странам были опубликованы в шестьдесят восьмом статистическом обзоре мировой энергетики в июне 2019 года.

В работе проанализирована динамика экспорта нефти и природного газа в двух состояниях (сжиженном и газообразном) с 2001 по 2018 год, динамика объемов добычи сырья и внутреннего потребления для стран мира с 2001 по 2018 год, выявлены мировые лидеры экспорта сырья, а также на основе индексов центральности выделены страны с наибольшим влиянием в сети для 2004, 2009, 2014 и 2018 годов.



Также отметим, что изучение международных конфликтов с помощью сетевого анализа позволяет оценить вовлеченность государств в международные конфликты в разные периоды времени, а также оценить влияние третьих стран на конфликт с помощью построения сетевой модели конфликтов. В модели учитывается интенсивность международного конфликта, интенсивность внутренних конфликтов в странах – участниках международных конфликтов, а также взаимодействие, основанное на поддержке основных участников конфликта.

Для оценки вовлеченности в конфликты используются не только классические индексы центральности, но и новые, учитывающие параметры вершин, групповое влияние вершин, а также позволяющие оценивать дальние и ближние взаимодействия вершин, что позволяет учитывать не только прямые, но и косвенные взаимосвязи [87], в том числе в сети конфликтов.

Для создания модели использованы наборы данных Упсальской программы данных о конфликтах (UCDP) Департамента исследований мира и конфликтов Упсальского университета и Центра изучения гражданской войны Международного института исследования мира в Осло (PRIO): «UCDP/PRIO Armed Conflict» [89] и «UCDP Battle-Related Deaths» [90], которые содержат информацию о вооруженных конфликтах с 1946 года по 2019 год.

В работе учитываются три типа конфликтов: межгосударственный вооруженный конфликт, внутренний вооруженный конфликт и интернационализованный внутренний вооруженный конфликт. Для оценки интенсивности конфликта используется информация о количестве погибших в результате боевых действий, которая преобразуется в шкалы. Модель применена для оценки вовлеченности государств в международные конфликты в разные периоды времени после Второй мировой войны.

Другим важным аспектом анализа сетей является то, что в рамках одной сети объекты могут взаимодействовать на различных уровнях. Например, сеть может состоять из разных слоёв (многослойные сети) [91] или меняться со вре-

менем [92]. Естественно, можно анализировать каждый слой или срез во времени независимо друг от друга, но также можно исследовать такие сложные сетевые структуры как единое целое.

Одним из примеров многослойной временной сети является сеть международной торговли продовольствием. Каждый слой такой сети представляет международную торговлю по конкретному виду товара, которая изменяется в динамике. Сеть международной торговли изучалась в работе [93] в контексте сокращения числа слоёв. Структурные свойства данной сетевой структуры изучались в работах [94], [95]. Влиятельность стран в контексте экспорта/импорта продовольствия изучались в работах [96], [97].

В данном исследовании нашей целью является анализ схожести слоёв сети торговли по различным видам продовольствия, а также выявления трендов в динамике за 20-летний период времени. Исследования корреляции слоёв сети является важным инструментом для анализа взаимозаменяемости продуктов, в то время как анализ динамики сети поможет нам выявить основные изменения в процессе торговли продовольствием.

Для анализа схожести сетей могут быть применены различные меры схожести вершин в графах [98], [99]. С их помощью можно сравнить одну и ту же вершину в разных слоях сетевой структуры. Также может быть использована идея ассортативного смешивания, основанная на корреляции степеней вершин [81], [100]. Анализ схожести сетей на основе сравнения подграфов изучен в работах [101], [102]. Спектральное разложение для сравнения сетей используется в работе [103]. Подход, основанный на энтропии фон Неймана, применён в работе [93] с целью уменьшения числа слоёв в сетевой структуре торговли продовольствием. Многокритериальный подход для сравнения сетей описан в работе [104].

Большинство из существующих подходов не учитывают такие аспекты, как вес и направления связей, возможность формирования групп вершин, структуру сообществ и другие важные элементы сложной сетевой структуры. Более того, нигде не учитывается важность вершин и рёбер в графах, в то

время как в реальных сетях важную роль играет изначальная влияние вершин. В данном исследовании мы учли вышеизложенные факты и адаптировали существующие классические сетевые метрики, а также разработали новый подход для оценки схожести слоёв сетевой структуры.

Ещё одним примером применения сетевого анализа данных является гравитационная модель торговли. Гравитационная модель торговли является одним из самых эффективных средств анализа торговых потоков. Ее классический вид предложен в работе [105]. Не смотря на высокую эффективность эмпирических оценок, довольно долгое время для данной модели не существовало строгого теоретического обоснования. Главной проблемой являлась невозможность объяснения мультипликативного эффекта вхождения экономических размеров и расстояния между странами в данную модель. Первой успешной попыткой обоснования модели на основе микроэкономической теории сделана Андерсоном [106]. В дальнейшем, гравитационная модель была расширена за счет добавления многостороннего сопротивления торговле [107]. Данный термин подразумевает под собой агрегированные барьеры в торговле для данной страны со всеми торговыми партнерами. Проблема использования расширенной модели в том, что при эконометрической оценке возникает смещение, вызванное зависимостью многостороннего сопротивления торговле от двусторонних барьеров. По этой причине было предложено несколько способов аппроксимировать многосторонние барьеры за счет использования симметричных барьеров и прокси-переменных [108] или использовать ряды Тейлора [109].

В последнее десятилетие стали популярны работы, предлагающие применить методы сетевого анализа для оценки многосторонних барьеров. Примеры использования индексов центральности для описания международной торговли можно найти в [110], [111]. Чаще всего индексы центральности добавляются в модель как признаки в регрессии. Однако в нашем исследовании, результаты которого приведены в подразделе 2.5, предлагается рассматривать также и отношения индексов центральности стран экспортеров и импортеров.

Такой подход позволяет лучше оценить важность данной страны для мировой торговли в сравнении с ее торговыми партнерами. Более того, рассматривая разницу между центральностями торговых партнеров можно получить больше информации о возможных двусторонних барьерах между ними.

Еще одним направлением исследования в сетевом анализе данных, проводимых в МЦАВР, стало оцениванию центральностей многослойной сети методами теории функций доверия. Отличительной особенностью этого подхода является то, что в нём агрегируются не центральности каждого слоя, а так называемые функции доверия локальных описаний сети в каждом слое. После чего вычисляются центральности по той информации, которую предоставляет новая (агрегированная) функция доверия. В этом случае агрегируется в каком-то смысле «первичная» информация о слоях. Поэтому данный подход представляется одним из перспективных направлений исследования.

## 2.2 Разработка обобщенного индекса LRIC, учитывающего влияние в сетевых структурах

### 2.2.1 Описание модели

Предлагается модель оценки влияния в сетевых структурах с учетом взаимозависимости вершин. Проиллюстрируем на рисунке 2.1 основную идею того, в какой ситуации наличие ребра между двумя вершинами может приводить к зависимости друг относительно друга.

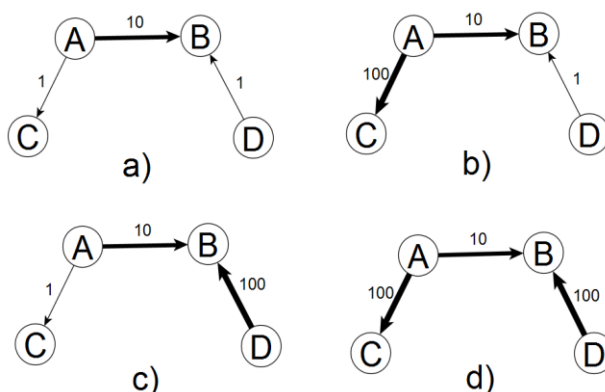


Рисунок 2.1 – Взаимозависимость вершин в сетевых структурах

Рассмотрим связь между вершинами А и В и допустим, что величина потока между вершинами соответствует объему торговли. Как показано на рисунке 2.1а, вершина А является основным экспортером для вершины В, поэтому можно сделать вывод, что вершина А влияет на вершину В. Тем не менее, вершина В также является крупнейшим импортером для вершины А, поэтому можно сделать вывод, что приостановка торговли между данными вершинами может экономически повлиять на вершину А. Таким образом, вершины А и В зависят друг от друга.

Рассмотрим теперь рисунок 2.1b. Как и ранее, вершина А является крупнейшим экспортером для страны В, но теперь вершина В не является крупнейшим импортером для вершины А. В такой ситуации будем говорить, что наличие связи между А и В приводит к зависимости В от А. Наоборот, рисунок 2.1с соответствует ситуации, когда вершина А зависит от вершины В: А не является крупнейшим экспортером для В, но В является крупнейшим импортером для А. Наконец, рисунок 2.1d соответствует ситуации, когда вершины А и В не зависят друг от друга, поскольку поток торговли между ними не является существенным для данных вершин. Отметим, что существующие методы не учитывают приведенную ситуацию.

Таким образом, возникает необходимость в преобразовании исходной сети с целью определения того, какую роль играет каждое ребро при оценке взаимозависимости вершин.

В рамках данной работы предлагается модель преобразования сети, которая используется в разработанном ранее индексе дальних взаимодействий LRIC [83], [84]. Согласно данной модели, каждая вершина имеет некоторый порог, при превышении которого другие вершины влияют на нее. Интенсивность влияния индекса LRIC оценивается с учетом групповых взаимодействий. В нашей ситуации, поскольку важность каждого ребра рассматривается одновременно для двух сторон, а каждая вершина может иметь как исходящие, так и входящие ребра, то будем считать, что каждая вершина имеет 2 пороговых значения  $q_i^{in}$  и  $q_i^{out}$ . Величина  $q_i^{in}$  соответствует пороговому значению

вершины  $i$ , при превышении которого входящие связи влияют на вершину  $i$  (влияние на вершину как импортера). Аналогично,  $q_i^{out}$  соответствует пороговому значению вершины  $i$ , при превышении которого исходящие связи влияют на вершину  $i$  (влияние на вершину как экспортера). Таким образом, индекс LRIC применяется к исходной сети 2 раза, а результаты можно представить в виде многослойной сети (рисунок 2.2), где один слой  $C^{out} = [c_{ij}^{out}]_{n \times n}$  показывает прямое влияние вершин как экспортеров (influence level), а другой  $C^{in} = [c_{ij}^{in}]_{n \times n}$  – их прямую зависимость как импортеров (dependence level).

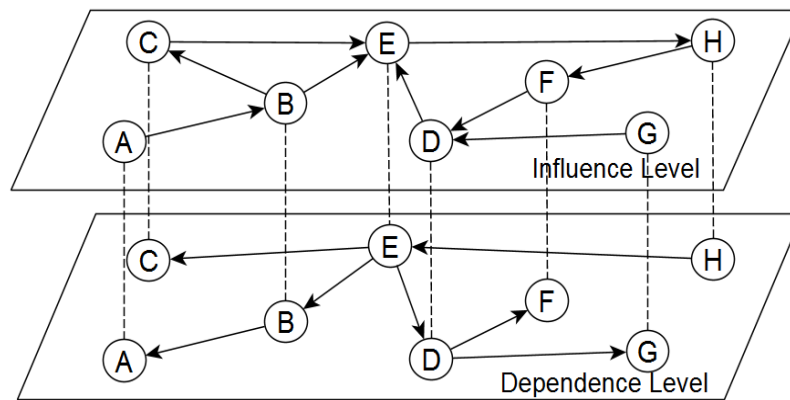


Рисунок 2.2 – Представление взаимозависимости вершин в виде многослойной сети

Отметим, что матрицы  $C^{in}$  и  $C^{out}$  показывают прямое влияние между вершинами. Тем не менее, поскольку вершины сети могут также влиять друг на друга через какие-то промежуточные вершины, возникает необходимость в оценке непрямого влияния вершин друг на друга, которое описывается матрицей  $\tilde{C} = [\tilde{c}_{ij}]_{n \times n}$ . В рамках данного исследования были предложены следующие модели агрегирования:

*Модель 1:* построение матрицы  $\tilde{C}$  на основе оценки максимального влияния вершин в каждой сети  $C^{in}$  и  $C^{out}$ . Согласно данной модели, каждый слой преобразованной сети (influence/dependence level) рассматривается отдельно и с помощью индекса LRIC происходит оценка непрямого влияния (матрицы  $\tilde{C}^{in}$  и  $\tilde{C}^{out}$ ). Величина  $\tilde{c}_{ij}^{in}$  показывает, насколько сильно вершина  $i$  зависит от

вершины  $j$ , а  $\tilde{c}_{ij}^{out}$  - насколько сильно вершина  $i$  влияет на вершину  $j$ . Итоговое влияние рассчитывается как  $\tilde{c}_{ij} = \tilde{c}_{ij}^{in} - \tilde{c}_{ij}^{out}$ .

*Модель 2:* преобразование сети прямых влияний и нахождение максимального пути между вершинами. Поскольку  $c_{ij}^{in}$  и  $c_{ij}^{out}$  характеризуют одно и то же ребро сети, можно преобразовать исходную сеть как  $c_{ij} = c_{ij}^{in} - c_{ij}^{out}$ . В таком случае получившиеся ребро показывает, для какой из вершины исходная связь в сети является более важной. Таким образом, матрица  $C = [c_{ij}]_{n \times n}$  показывает прямое влияние вершин друг на друга. На последующем этапе матрица прямых влияний  $C$  преобразуется в матрицу непрямых влияний  $\tilde{C}$  в соответствии с индексом дальних взаимодействий LRIC.

*Модель 3:* одновременный учет информации многослойной сети. Согласно данной модели итоговое влияние между вершинами определяется с помощью нахождения такого пути  $P_{i \rightarrow j}$ , на котором наблюдается наибольшая разница между влиянием и зависимостью вершин  $i$  и  $j$ , т.е.

$$\tilde{c}_{ij} = \max_{P_{i \rightarrow j}} (f^{in}(P_{i \rightarrow j}) - f^{out}(P_{i \rightarrow j})).$$

Построенная сеть непрямых влияний  $\tilde{C}$  может быть преобразована в новый индекс центральности в соответствии с методологией индекса дальних взаимодействий [83], [84].

### 2.2.2 Применение обобщенного индекса LRIC к анализу сети международной торговли

Рассмотрим применение предложенной модели к сети международной торговли. В качестве исходных данных была использована информация об общей торговле между странами за 2017 год, предоставленная Организацией экономического сотрудничества и развития (ОЭСР/OECD) [112]. Рассматриваемая сеть является направленной и взвешенной, общее число вершин – 198, общее число ребер – 27264.

На первом этапе произведем расчет классических индексов центральности (Таблица 2.1).

Таблица 2.1 – Ранжирование вершин сети международной торговли с помощью классических индексов центральности (2017)

Вершины сети	Индексы центральности				
	PageRank	Eigenvector	Hubs	Authorities	Weighted Out-degree
США	1	17	5	1	2
Китай	2	1	1	3	1
Германия	3	16	4	5	3
Франция	4	2	10	10	6
Великобритания	5	3	12	9	10
Япония	6	21	6	4	4
Канада	7	19	3	7	12
Гонконг	11	40	8	2	7
Мексика	12	13	2	6	11
Республика Корея	13	10	7	8	5
Чехия	28	4	33	30	29

Как показано в Таблице 2.1, наиболее важными вершинами в сети международной торговли являются США, Китай, Германия. Последующие места занимают такие страны, как Франция, Великобритания, Канада, Мексика, Гонконг (Китай), а также другие страны. Отметим, что данные индексы центральности не учитывают индивидуальные атрибуты вершин (стран), а также возможность их группового влияния.

Перед тем, как перейти к расчету обобщенного индекса LRIC, необходимо определить порог влияния на каждую страну. Поскольку исходные данные соответствуют потокам общей торговли, то в рамках данного исследования было решено использовать информацию о валовом внутреннем продукте (ВВП) стран для оценки порога влияния. Другими словами, порог рассчитывался как некоторая доля  $\lambda$  от ВВП страны  $i$  ( $GDP_i$ ), т.е.

$$q_i^{out} = q_i^{in} = \lambda \cdot GDP_i.$$

В рамках данной работы было рассмотрено 2 уровня порога: низкий ( $\lambda = 0.1$ ) и высокий ( $\lambda = 0.3$ ). Результаты расчетов приведены в Таблице 2.2.

Таблица 2.2 – Ранжирование вершин сети международной торговли с помощью обобщенных индексов LRIC (2017)

Вершины сети	$\lambda = 0.1$			$\lambda = 0.3$		
	Модель 1	Модель 2	Модель 3	Модель 1	Модель 2	Модель 3
Китай	1	1	1	3	2	2
США	2	2	2	1	1	1



## Продолжение таблицы 2.2

Германия	3	3	3	2	3	3
Франция	4	6	4	4	4	4
Великобритания	5	7	6	5	7	7
Япония	7	5	5	6	5	5
Италия	8	4	7	11	8	10

Как показано в Таблице 2.2, наиболее важными вершинами в сети международной торговли являются Китай и США. Германия, Франция, Великобритания, Япония и Италия занимают последующие места. Если сравнивать различные модели между собой, то мы можем сделать вывод, что модели имеют высокую корреляцию друг с другом (более 0.9). Отметим, что результаты предложенных индексов отличаются от классических индексов центральности, что объясняется учетом индивидуальных характеристик стран (ВВП), возможностью группового влияния, а также учетом того, насколько важен каждый торговый поток сети, как для экспортера, так и для импортера. Более того, предложенная модель позволяет проанализировать попарное влияние между странами. В частности, было выявлено, что в соответствии с предложенной моделью Китай влияет в одиночку на 46 стран, а США – на 36 стран.

Более подробная информация о проведенном исследовании приведена в [113].

### 2.2.3 Применение индекса LRIC к анализу сети взаимодействий террористических групп

Предложенные модели оценки дальних взаимодействий могут быть также применены при анализе взаимодействий террористических групп. Поскольку террористические организации могут влиять на политические процессы в определенной стране, регионе или во всем мире, они становятся важными участниками мировых взаимоотношений. Так как данные организации взаимодействуют друг с другом, можно построить всемирную сеть, отражающую отношения между ними, и провести ее детальный анализ.

В рамках данного исследования был произведен сетевой анализ взаимодействия террористических групп с 2001 по 2018 годы. В качестве исходных

данных была взята база данных [114]. Была построена международная сеть террористических групп и посчитаны индексы влияния групп в динамике. Вершинами сети являются различные террористические организации, а ребра сети отражают информацию об общем числе совершенных совместных террористических атак и числе жертв (пример сети приведен на рисунке 2.3).

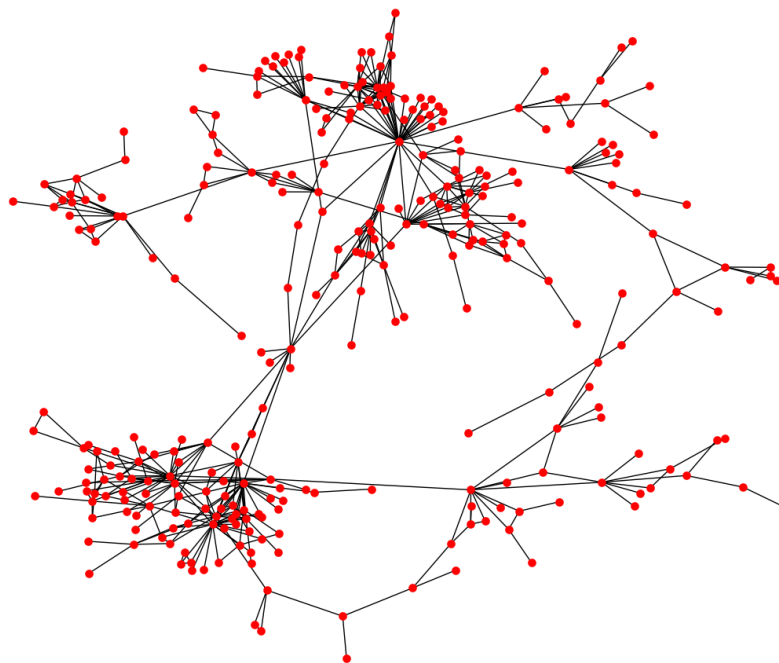


Рисунок 2.3 – Взаимодействие террористических групп (2001-2018)

Проанализировав частоту террористических атак, организованных всеми группами сети (в одиночку или в сотрудничестве с другими группами), можно также обнаружить, что террористическая активность менялась за годы наблюдения. Более того, некоторые организации действовали в начале 2000-х годов и не осуществляли свою деятельность в 2010-х годах. В связи с этим мы поделили 2001-2018 годы на три различных однородных периода: 2001-2007, 2008-2011, 2012-2018. В рамках каждого периода были проанализированы наиболее важные участники сети с использованием, как классических индексов центральности, так и индексов ближних и дальних взаимодействий SRIC/LRIC.

В Таблицах 2.3, 2.4 приведены полученные результаты для крупнейшей компоненты связности за 2012-2018 годы. Отметим, что порог влияния на вершину определялся как некоторый % от общего числа атак или жертв.

Таблица 2.3 – Индексы SRIC/LRIC, 2012-2018 (сеть по общему числу атак, ТОП-3)

Название группы	Центральность (ранжирование)					
	5%		30%		60%	
	SRIC	LRIC	SRIC	LRIC	SRIC	LRIC
Taliban	0.071 (1)	0.034 (8)	0.042 (7)	0.03 (10)	0.003 (50)	0.002 (47)
Islamic State of Iraq and the Levant (ISIL)	0.07 (2)	0.038 (5)	0.049 (6)	0.032 (8)	0.027 (12)	0.02 (13)
Tehrik-i-Taliban Pakistan (TTP)	0.054 (3)	0.052 (1)	0.053 (3)	0.04 (6)	0.023 (14)	0.016 (16)
Lashkar-e-Jhangvi	0.05 (4)	0.043 (3)	0.019 (18)	0.024 (12)	0.031 (9)	0.026 (11)
Lashkar-e-Taiba (LeT)	0.026 (6)	0.027 (9)	0.052 (4)	0.056 (3)	0.014 (22)	0.004 (40)
Al-Nusrah Front	0.022 (10)	0.045 (2)	0.078 (1)	0.076 (1)	0.097 (1)	0.078 (3)
Hizbul Mujahideen (HM)	0.014 (19)	0.019 (12)	0.055 (2)	0.063 (2)	0.022 (15)	0.013 (19)
Badr Brigades	0.0086 (31)	0.001 (144)	0.025 (11)	0.0226 (14)	0.084 (2)	0.118 (1)
Asa'ib Ahl al-Haq	0.0085 (32)	0.002 (116)	0.024 (12)	0.0027 (13)	0.077 (3)	0.114 (2)

Таблица 2.4 – Индексы SRIC/LRIC, 2012-2018 (сеть по общему числу жертв, ТОП-3)

Название группы	Центральность (ранжирование)					
	5%		30%		60%	
	SRIC	LRIC	SRIC	LRIC	SRIC	LRIC
Haqqani Network	0.29 (1)	0.064 (4)	0.05 (6)	0.051 (6)	0.06 (4)	0.05 (4)
Khorasan Chapter of the Islamic State	0.11 (2)	0.093 (1)	0.03 (11)	0.026 (12)	0.03 (9)	0.028 (8)
Lashkar-e-Taiba (LeT)	0.1 (3)	0.057 (6)	0.04 (7)	0.052 (5)	0.04 (7)	0.049 (5)
Tehrik-i-Taliban Pakistan (TTP)	0.05 (4)	0.088 (2)	0.12 (2)	0.055 (4)	0.027 (11)	0.017 (16)
Taliban	0.04 (5)	0.08 (3)	0.13 (1)	0.11 (2)	0.18 (1)	0.15 (1)
Al-Nusrah Front	0.03 (9)	0.039 (7)	0.09 (3)	0.12 (1)	0.16 (2)	0.14 (2)
Ahrar al-Sham	0.02 (10)	0.033 (8)	0.08 (4)	0.1 (3)	0.07 (3)	0.08 (3)

Как показано в Таблицах 2.3, 2.4, с помощью индексов ближних и дальних взаимодействий (SRIC, LRI) были определены наиболее влиятельные террористические организации в сети взаимоотношений между ними. Несмотря на то, что индексы SRIC и LRIC были рассчитаны для различных порогов влияния, результаты оказались практически идентичными. В целом, индексы SRIC и LRIC показали похожие результаты с классическими индексами центральности, но они также позволили выявить новые влиятельные организации. Результаты SRIC и LRIC для невзвешенной сети и сети, построенной по числу

совершенных атак, в основном совпадают, в то время как сеть, построенная по общему числу жертв, выявила новые влиятельные группы. Таким образом, используя сетевой подход можно определить мощные и влиятельные террористические группы, которые на первый взгляд могут показаться неочевидными.

Вся работа была также проведена на языке программирования Python, расчеты индексов центральности были осуществлены с помощью пакетов 'networkx' и 'slric'. Более подробная информация о проведенном исследовании приведена в [115].

## **2.3 Разработка и применение индексов центральности, учитывающих индивидуальные и групповые влияния**

### **2.3.1 Математическая модель новых индексов центральности**

Пусть  $G^0 = (V, W^0)$  является взвешенным направленным графом, где  $V$  – набор вершин,  $V = n$ ,  $W^0$  – набор дуг с весами  $w_{ji}^0$ . Для каждой вершины  $i \in V$  определена квота  $q_i$ . Максимальное количество дуг, одновременно влияющих на некоторую определенную вершину, составляет  $k$ . Тогда на основе этих предпосылок можно перейти к описанию обобщенного индекса центральности, который состоит из трех индексов: (1) CI (от англ. «copeland in-degree centrality index»), (2) BI (от англ. «bundle influence in-degree centrality index»), (3) PI (от англ. «pivotal influence in-degree centrality index»).

CI – это взвешенный по степени индекс центральности, который суммирует веса входящих ребер:  $CI^0(i) = \sum_j w_{ji}^0$ .

BI подсчитывает общее число критических наборов, для которых суммарные веса входящих ребер больше или равны заданной квоте:

$$BI_i^0(S) = \begin{cases} 1, & \text{if } \sum_{j \in S} w_{ji}^0 \geq q_i, \\ 0, & \text{else.} \end{cases}$$

PI оценивает общее число ключевых вершин в сети:  $PI^0(i) = \sum_S PI_i^0(S)$ , где  $S \subseteq V \setminus \{i\}$ ,  $|S| \leq k$ ,  $\forall j \in S, w_{ji}^0 \neq 0$ .

Более подробное описание индексов можно найти в работе [87].

### 2.3.2 Анализ сети международной миграции

Одним из методов сетевого анализа миграции является оценивание влияния стран с помощью индексов центральности. В данной работе индексы, предложенные в [87], применены к проблеме международной миграции. Данные индексы имеют ряд преимуществ по сравнению с классическими индексами, в том числе, в области международной миграции.

Для анализа международной миграции в работе были использованы данные Евростата и ОЭСР [116], [117] по потокам миграции между странами мира. В базах данных страна происхождения мигрантов определяется на основе гражданства, места рождения или проживания в зависимости от миграционного учета в принимающих странах. На основе этих данных строится сеть миграции, в которой вершинами являются страны, а ребрами – миграционные потоки между ними. Основная задача состоит в оценке влияния стран в сети. Индексы, используемые в работе, учитывают как прямое влияние в сети, так и групповое влияние, и не прямые взаимодействия в сети.

Предлагаемые индексы центральности также учитывают индивидуальные характеристики узлов сети, а именно демографическую характеристику – общую численность населения. Общая численность населения этих стран имеет важное значение с точки зрения процесса международной миграции. Предположим, что существует миграционный поток в 50 тысяч мигрантов, прибывающих в страну А с населением 100 миллионов человек, и такой же миграционный поток, прибывающий в страну Б с населением 500 тысяч человек. Этот поток будет составлять 0,05% от населения страны А, однако он будет достигать 10% от населения страны Б, что может быть критично для населения страны Б и почти незаметно для населения страны А. Таким образом, индивидуальные параметры узлов, в данном случае, населения очень важны при оценке влияния стран в сети, в частности, в процессе международной миграции.

В целом влияние международной миграции в данной работе моделируется следующим образом. Новые индексы центральности оценивают групповое влияние стран с разных точек зрения. Bundle Influence (BI) соответствует групповому влиянию критически больших наборов стран, имеющих миграционный поток в страну назначения. Pivotal Influence (PI) оценивает влияние ключевой страны внутри каждой из таких групп. Индекс Коупленда (CI) соответствует притоку мигрантов в страну. Общее влияние – Total Influence (TI) оценивается как линейная комбинация трех показателей. Индивидуальные характеристики стран учитываются с помощью квоты, которая соответствует процентной доле от населения принимающей страны.

Результаты оценки индексов BI и PI для топ-10 стран представлены при различных значениях критической группы и квоты.

Во-первых, индекс BI зависит от размера группы. С увеличением размера от 3 до 4 мы наблюдаем смену мест между 9-й и 10-й странами: Япония переместилась на 9-е место, а Австрия стала 10-й (Таблица 2.5). Дальнейшее увеличение размера группы приводит к изменению верхних мест, то есть США становятся 1-й страной, а Германия – 2-й в рейтинге. Это означает, что позиции этих стран (США и Японии) чувствительны к размерам групп стран, которые направляют в них мигрантов. Для Японии групповое влияние становится более значительным с увеличением размера группы с 3 до 4, в то время как для Соединенных Штатов увеличение с 4 до 5 стран в группе оказывает влияние на позицию в рейтинге.

Таблица 2.5 – Топ-10 стран по индексу Bundle Influence в 2015 для размеров групп 3, 4 и 5

BI group size 3	BI group size 4	BI group size 5
Germany	Germany	United States
United States	United States	Germany
United Kingdom	United Kingdom	United Kingdom
Spain	Spain	Spain
France	France	France
Canada	Canada	Canada
Italy	Italy	Italy
Australia	Australia	Australia
Austria	Japan	Japan
Japan	Austria	Austria

Во-вторых, индекс Pivotal Influence еще больше зависит от размера группы (Таблица 2.6). PI указывает на влияние ключевой страны в группе на принимающую страну. В Таблице 2.6 представлен рейтинг стран, которые в наибольшей степени подвержены влиянию ключевой страны в группе. Германия сменила место в рейтинге с 1-го для группы размера 3 на 3-е для группы размера 5, в то время как Великобритания и Соединенные Штаты переместились выше. Кроме того, Япония переместилась с 8-го на 5-е место, а Мексика оказалась последней в рейтинге.

Таблица 2.6 – Топ-10 стран по индексу Pivotal Influence в 2015 для размеров групп 3, 4 и 5

PI group size 3	PI group size 4	PI group size 5
Germany	United States	United States
United States	Germany	United Kingdom
United Kingdom	United Kingdom	Germany
France	France	France
Italy	Italy	Japan
Spain	Japan	Italy
Canada	Spain	Spain
Japan	Canada	Canada
Republic of Korea	Republic of Korea	Republic of Korea
Poland	Poland	Mexico

Теперь перейдем к изменениям во влиянии стран, связанными с увеличением квоты для размера группы 3. Таблица 2.7 показывает, что Германия является страной, наиболее устойчивой к изменениям размера квоты, однако существует вариативность в рейтингах других стран. Наиболее заметным изменением является переход Австрии с 8-го места на 2-е. В 2015 году наибольший миграционный поток в Австрию был из Сирии (22188 мигрантов), другими странами, отправляющими мигрантов в Австрию, были Афганистан, Германия, Румыния, Венгрия. Основные изменения индексов VI и PI в связи с увеличением квоты очень схожи.

Таблица 2.7 – Топ-10 стран по Bundle Influence и Pivotal Influence в 2015 году для размера группы 3 и различных размеров квот (q)

BI q=0,01%	BI q=0,05%	BI q=0,1%	PI q=0,01%	PI q=0,05%	PI q=0,1%
Germany	Germany	Germany	Germany	Germany	Germany
United States	United Kingdom	Austria	United States	United Kingdom	United Kingdom
United Kingdom	United States	United Kingdom	United Kingdom	United States	Austria
Spain	Austria	Switzerland	France	Spain	Switzerland
France	Spain	Canada	Italy	Sweden	Belgium
Canada	Switzerland	Australia	Spain	Canada	Australia
Italy	Australia	Belgium	Canada	Australia	Canada
Australia	Canada	New Zealand	Japan	Switzerland	Sweden
Austria	Sweden	Sweden	Republic of Korea	Austria	New Zealand
Japan	Belgium	Netherlands	Poland	Netherlands	Denmark

До сих пор мы рассматривали только индексы Bundle Influence и Pivotal Influence при разных размерах групп и величине квоты. Однако существует индекс суммарного влияния, представляющий собой линейную комбинацию индексов CI, BI и PI. Присвоение этим индексам различных весов приводит к изменению интерпретации суммарного индекса влияния. Более высокий вес для влияния Коупленда относится к интерпретации, более близкой к притоку мигрантов. Если BI имеет большее значение в общем влиянии, то интерпретация TI смещается в сторону группового влияния на страну. Наконец, большее значение PI приводит к более высокому влиянию ключевых стран в группе на принимающую страну. Далее мы рассмотрим случаи с различными весами в TI для фиксированного размера группы, равного трем, и квоты, равной 0,01% от населения принимающей страны.

Таблица 2.8 – Топ-10 стран по индексу Total Influence в 2015 г. с различными весами

TI_40_40_20	TI_50_30_20	TI_60_30_10	TI_70_20_10	TI_80_10_10
Germany	Germany	Germany	Germany	Germany
United States	United States	United States	United States	United States
United Kingdom	United Kingdom	United Kingdom	United Kingdom	United Kingdom
France	France	Spain	Spain	Japan
Spain	Spain	France	Japan	Republic of Korea
Canada	Italy	Japan	Republic of Korea	Spain
Italy	Canada	Canada	France	France
Japan	Japan	Italy	Italy	Italy
Republic of Korea	Republic of Korea	Republic of Korea	Canada	Canada
Australia	Australia	Australia	Australia	Australia



При увеличении веса влияния Коупленда (притока мигрантов) в позициях топ-3 стран не происходит никаких изменений (Таблица 2.8). Однако Япония и Республика Корея поднимаются в рейтинге выше, в то время как Франция, Испания и Италия снижают свои позиции. Китай имеет самое высокое индивидуальное влияние (приток мигрантов в Японию и Республику Корея составляет 100571 и 177001 человек соответственно). Напротив, миграционный поток в европейские страны в рейтинге (Франция, Испания и Италия) более диверсифицирован, и поэтому с увеличением веса  $CI$  в общем влиянии влияние европейских стран уменьшается.

### 2.3.3 Анализ сети общей торговли

Индексы, описанные в работе [87], были рассчитаны для международной торговой сети на следующем временном промежутке: 2001, 2006, 2011 и 2016 годы. При этом отметим, что источником данных послужила база статистических данных ООН по торговле товарами. Поскольку в данных о международной торговле есть существенные расхождения между экспортными и импортными значениями, то они были предварительно обработаны для получения единого значения для одного и того же торгового потока для импортера и экспортера.

Для расчета индексов центральности было принято максимальное число стран, которые могут одновременно влиять на страну (число стран в критическом множестве):  $k = 5$ . Что касается квот  $q_i$ , то для международной торговли она принималась в 10% от общего объема импорта в страну за рассматриваемый год.

Далее продемонстрируем полученные результаты для топ-30 наиболее влиятельных стран (Таблица 2.9).

Таблица 2.9 – VI для топ-30 стран

Country	2001	2006	2011	2016
France	0.027	0.021	0.020	0.021
Spain	0.025	0.019	0.021	0.019
South Africa	0.023	0.020	0.018	0.018
Poland	0.007	0.020	0.018	0.018
Netherlands	0.019	0.019	0.019	0.018

Продолжение таблицы 2.9

Country	2001	2006	2011	2016
Thailand	0.019	0.017	0.016	0.018
Germany	0.027	0.021	0.018	0.018
Korea, Rep.	0.019	0.019	0.015	0.017
United Kingdom	0.024	0.021	0.017	0.017
United States	0.024	0.021	0.018	0.017
Belgium	0.015	0.013	0.015	0.017
Russian Federation	0.014	0.013	0.013	0.016
Canada	0.020	0.017	0.016	0.016
Czech Republic	0.017	0.017	0.016	0.016
Singapore	0.015	0.005	0.008	0.016
New Zealand	0.007	0.014	0.012	0.015
Pakistan	0.007	0.013	0.012	0.015
India	0.011	0.017	0.015	0.015
Turkey	0.011	0.014	0.015	0.014
Slovenia	0.015	0.015	0.014	0.014
Italy	0.019	0.019	0.015	0.014
Slovak Republic	0.010	0.013	0.012	0.014
Ireland	0.016	0.013	0.014	0.014
China	0.011	0.013	0.013	0.014
Sweden	0.013	0.013	0.011	0.014
Mexico	0.011	0.017	0.014	0.014
Australia	0.015	0.014	0.014	0.013
Finland	0.011	0.011	0.005	0.013
Brazil	0.013	0.013	0.014	0.013
Denmark	0.016	0.014	0.013	0.013

Можно видеть, что Франция, Испания, Южная Африка, Польша, Нидерланды, Таиланд, Германия, Корея, Великобритания и Соединенные Штаты входят в число лучших стран по индексу ВІ (Таблица 2.9).

Может показаться удивительным, что Польша и Таиланд входят в число ведущих стран, но причины этого следующие. Во-первых, индекс ВІ зависит от количества торговых партнеров. Если страна имеет большое число торговых партнеров, то вероятность формирования коалиции выше, а если страна имеет небольшое число торговых партнеров, то – ниже. Действительно, если мы более внимательно посмотрим на общее количество торговых партнеров, то увидим, что Таиланд занимает второе место в мире по количеству торговых партнеров, Польша-четвертое. Для сравнения: Китай занимает тридцать четвертое место по числу торговых партнеров.

Во-вторых, индекс ВІ зависит от значений торговых потоков. Если торговые потоки страны распределены равномерно, то риски группового влияния

торговых партнеров ниже; если торговые потоки распределены неравномерно, то риски формирования коалиции против данной страны выше. Действительно, если мы более внимательно посмотрим на значения торговых потоков, то увидим, что на долю топ-5 экспортеров Таиланда приходится 57% тайского импорта, а на долю топ-5 экспортеров Польши приходится 52% польского импорта. Для сравнения: на долю топ-5 экспортеров Китая приходилось 34% китайского импорта.

Такие страны, как Франция, Таиланд, Словацкая Республика, Корея, Польша, Сингапур, Мексика, Чехия, Новая Зеландия, Канада, Великобритания, Германия, имеют высокие значения по индексу PI (Таблица 2.10).

Таблица 2.10 – PI для топ-30 стран

Country	2001	2006	2011	2016
France	0.027	0.021	0.019	0.022
Thailand	0.022	0.018	0.021	0.022
Slovak Republic	0.013	0.019	0.017	0.020
Korea, Rep.	0.021	0.022	0.017	0.020
Poland	0.006	0.024	0.022	0.019
Singapore	0.018	0.004	0.009	0.019
Mexico	0.015	0.022	0.017	0.019
Czech Republic	0.021	0.022	0.021	0.019
New Zealand	0.008	0.017	0.015	0.018
Canada	0.028	0.020	0.017	0.018
United Kingdom	0.025	0.019	0.016	0.018
Germany	0.028	0.022	0.016	0.017
Slovenia	0.019	0.017	0.014	0.017
South Africa	0.025	0.019	0.018	0.017
Pakistan	0.006	0.012	0.013	0.017
Belgium	0.014	0.012	0.015	0.017
Spain	0.023	0.015	0.017	0.017
Ireland	0.022	0.017	0.018	0.017
United States	0.025	0.019	0.016	0.016
Netherlands	0.019	0.019	0.018	0.016
Sweden	0.015	0.015	0.013	0.016
Indonesia	0.018	0.014	0.015	0.016
Australia	0.019	0.018	0.016	0.015
Bahrain	0.001	0.001	0.005	0.014
Finland	0.011	0.011	0.005	0.014
Russian Federation	0.012	0.012	0.011	0.014
Austria	0.022	0.018	0.018	0.014
Japan	0.022	0.018	0.014	0.013
Brazil	0.012	0.012	0.013	0.013
Denmark	0.016	0.014	0.013	0.013

Опять же, может быть удивительно, что Таиланд, Словацкая Республика, Польша, Сингапур входят в число ведущих стран. Причины такого распределения стран аналогичны тем, которые были указаны для индекса ВІ.

Во-первых, индекс РІ учитывает общее количество торговых партнеров. Чем больше у страны торговых партнеров, тем выше риск иметь ключевую страну, и наоборот. Поскольку Таиланд, Польша и Словакия являются одними из крупнейших торговых партнеров в мире, этот риск отражается в индексе РІ.

Во-вторых, индекс РІ также зависит от значений торговых потоков. Чем выше значение торгового потока, тем больше вероятность того, что торговый партнер будет ключевым. Если мы посмотрим на статистику, то увидим, что Таиланд, Словацкая Республика, Польша, Сингапур имеют высококонцентрированный импорт из других стран.

Таблица 2.11 – Обобщенный индекс центральности для топ-30 стран

Country	y2001	y2006	y2011	y2016
United States	0.040	0.033	0.028	0.029
France	0.030	0.024	0.022	0.023
Germany	0.032	0.027	0.022	0.022
China	0.014	0.017	0.019	0.020
United Kingdom	0.028	0.023	0.019	0.020
Korea, Rep.	0.020	0.021	0.017	0.019
Thailand	0.019	0.016	0.018	0.019
Netherlands	0.021	0.021	0.020	0.019
Poland	0.007	0.021	0.019	0.018
Spain	0.024	0.018	0.019	0.018
Canada	0.025	0.020	0.017	0.018
Belgium	0.016	0.014	0.016	0.017
Singapore	0.016	0.006	0.010	0.017
South Africa	0.022	0.018	0.017	0.016
Mexico	0.014	0.019	0.016	0.016
Czech Republic	0.017	0.018	0.017	0.016
Slovak Republic	0.010	0.015	0.013	0.015
New Zealand	0.007	0.014	0.012	0.015
Russian Federation	0.013	0.012	0.013	0.015
Japan	0.022	0.019	0.016	0.015
Pakistan	0.006	0.011	0.011	0.014
Slovenia	0.016	0.015	0.013	0.014
Ireland	0.017	0.014	0.015	0.014
Sweden	0.013	0.013	0.012	0.014
Australia	0.016	0.015	0.014	0.014
Italy	0.020	0.020	0.016	0.014
India	0.010	0.015	0.015	0.014
Indonesia	0.015	0.012	0.013	0.013
Turkey	0.010	0.013	0.014	0.013
Brazil	0.012	0.012	0.013	0.013

Общий индекс влияния для топ-30 стран выглядит следующим образом (Таблица 2.11). Заметим, что этот показатель позволяет выявить новые характеристики межстранового взаимодействия, которые еще не рассматривались в сетевом анализе. В частности, данный индекс выделяет в качестве наиболее главных те страны, которые являются наиболее крупными импортерами и экспортерами.

#### 2.3.4 Сетевой анализ экспорта и импорта нефти и газа

В работе была проанализирована динамика объемов поставок нефти и газа среди экономических содружеств, были рассмотрены такие как ЕС, Андский пакт, Азиатско-Тихоокеанское содружество, МЕРКОСУР – объединение стран южной Америки, организация экономического сотрудничества и развития (ОЭСР), ОПЕК, Североамериканское соглашение о свободной торговле (НАФТА), Европейская ассоциация свободной торговли (ЕАСТ). Для каждого из содружеств подготовлен подробный анализ каждой страны, входящей в содружество, а также указаны причины, почему та или иная страна сокращала или наращивала объемы экспорта/импорта – было ли это связано с изменением спроса, или пришлось изменить объемы добычи. Итоги по динамике экспорта нефти и газа в двух состояниях за 2001 – 2018 годы представлены в сводных графиках для развитых, развивающихся стран и стран с переходной экономикой. На момент работы с данными статистика по экспорту была представлена не у всех стран. Была использована классификация ООН 2019 года [118]. Соотношение стран по группам было произведено по реальному ВВП в ценах 2012 года в долларах США (рисунок 2.4).

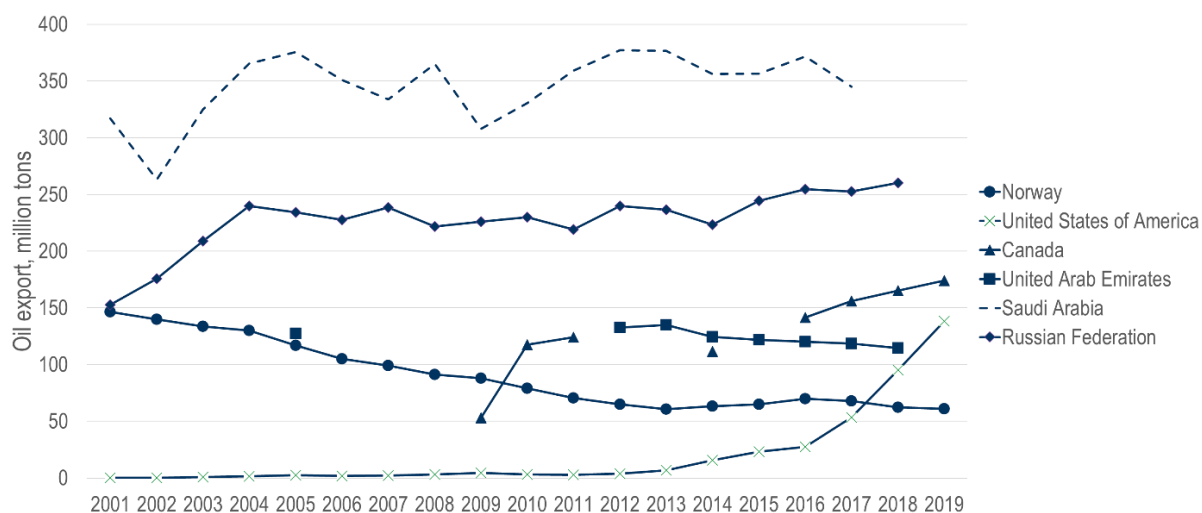


Рисунок 2.4 – Динамика экспорта нефти странами-лидерами в 2001 – 2019 гг., млн. тонн. Источник: данные организации ИТС [<http://www.intracen.org/>], расчеты автора

Мировой лидер по экспорту нефти – Саудовская Аравия, максимальное количество экспортируемой нефти наблюдалось в 2005, 2012, 2016 годах – 375, 377 и 371 млн тонн в год соответственно. Нефть из Саудовской Аравии отправляется в страны Европы, Азии. По объемам экспорта наблюдается восходящая тенденция, это, безусловно, связано с ростом добычи нефти – за период показатель вырос на 35%. Центровыми продавцами нефти среди развитых стран в настоящее время являются Канада и США. Канада опережает США по наращиванию темпов роста экспорта, но за 2016-2019 годы темпы роста немного снизились, к тому же США, изначально являясь импортерами нефти, благодаря сланцевой революции превратились в одних самых крупных экспортеров нефти в мире. Россия занимает второе место в мире по экспорту нефти, объем добычи за рассматриваемый период вырос на 60%, в 2018 году было добыто 563 млн. тонн нефти. Россия экспортирует нефть в страны Азии и Европы.

В экспорте природного газа в газообразном состоянии (рисунок 2.5) из развитых стран до 2009 года лидером была Канада, но с 2010 года до 2016-го показатель объема экспорта составлял в среднем 25 млн. тонн экспорта газа в год. США также экспортирует примерно на равном уровне с Канадой, но в последнее время темпы прироста увеличиваются и в ближайшем будущем

должны выйти на второе место после Норвегии, из-за того, что Канада сокращает экспорт. В России наблюдается резкий скачок в 2008 году, но последующее сильное сокращение, которое связано с кризисными явлениями в мировой экономике. Первое место в мире по объему экспорта газа на данный момент принадлежит России.

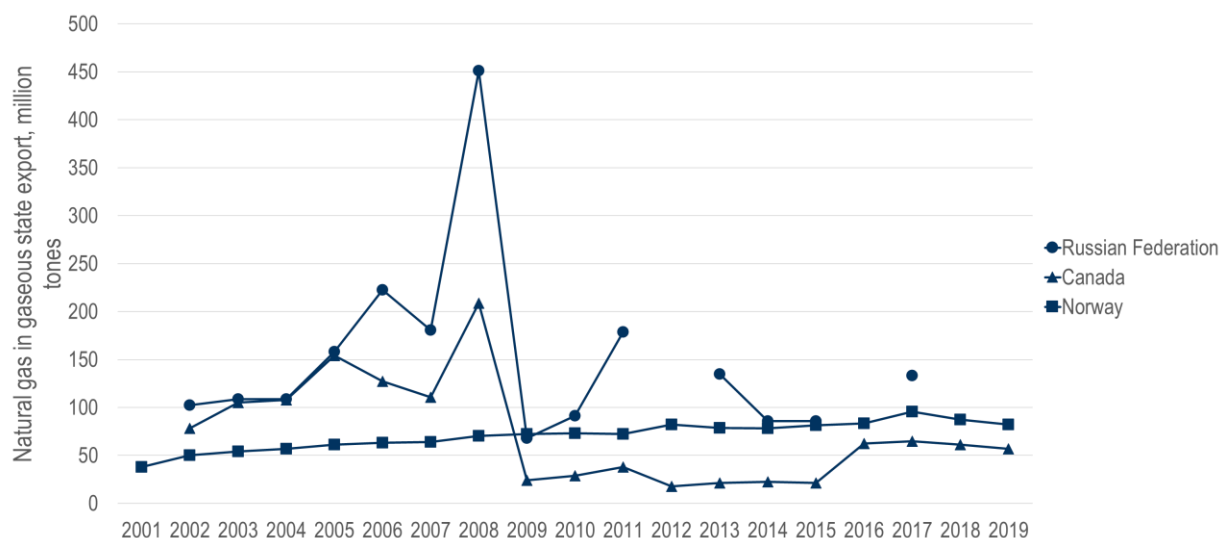


Рисунок 2.5 – Динамика экспорта природного газа в газообразном состоянии странами-лидерами в 2001 – 2019 гг., млн тонн. Источник: данные организации ИТС [<http://www.intracen.org/>], расчеты автора

Выделив лидеров по экспорту сжиженного газа из развивающихся стран (рисунок 2.6), можно заметить, что ведущее место занимает, безусловно, Катар. Он резко начал наращивать количество экспортируемого газа в 2007 году, а к 2015 году достиг максимума в рассматриваемый период – 120 млн тонн в год. Судя по восходящему тренду, Катар будет продолжать наращивать добычу и экспорт СПГ в ближайшее время. Тем не менее среди развитых стран, экспортирующих СПГ, на первое место выходит Австралия, в 2018 году она экспортировала почти 80 млн тонн. По прогнозам аналитиков, возможно, что к 2023 году экспорт США превысит австралийский.

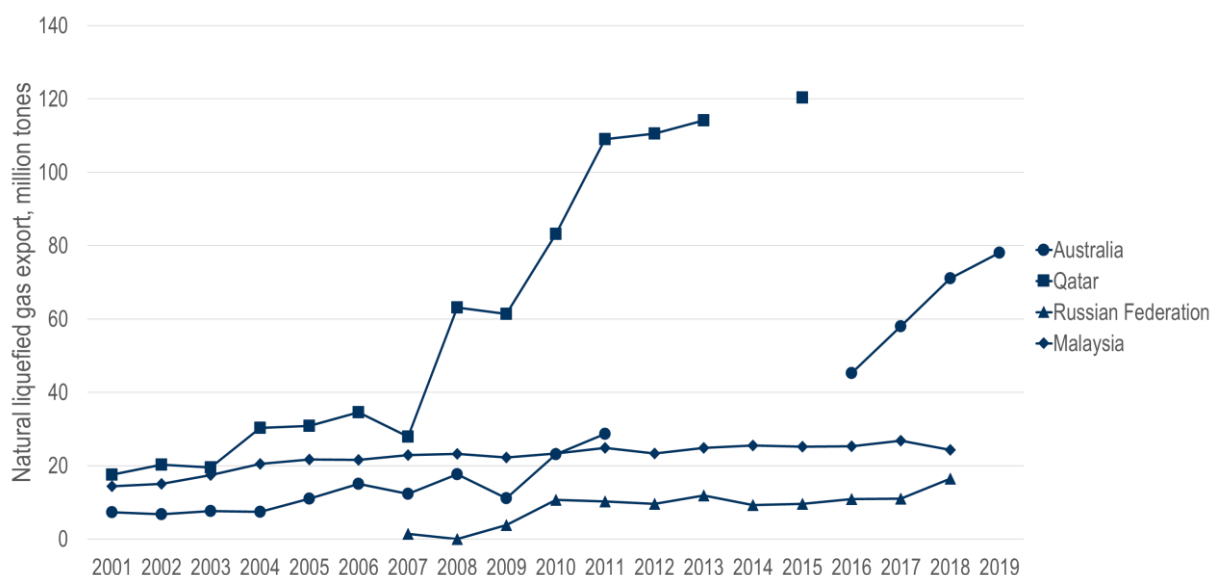


Рисунок 2.6 – Динамика экспорта сжиженного природного газа странами-лидерами в 2001 – 2019 гг., млн тонн. Источник: данные организации ИТС [<http://www.intracen.org/>], расчеты автора

В работе также были посчитаны основные индексы центральности InDegree, OutDegree, Degree, EigenVector, PageRank, Closeness Centrality, Betweenness Centrality. Ниже приведена часть рассчитанных значений для 2018 года для нефти, газа в газообразном и сжиженном состояниях.

Таблица 2.12 – Рейтинг стран, экспортирующих нефть, по индексам центральности на 2018 год

Country	InDeg	OutDeg	Deg	Clos	BetwCentr	PageRank	EigenVec
United States of America	1	2	1	8	1	3	4
China	2	23	5	39	10	7	3
Korea, Republic of	3	44	12	44	44	2	1
Netherlands	4	45	16	45	45	4	7
Singapore	5	46	18	46	46	6	2
Russian Federation	81	1	2	9	19	86	81
Kazakhstan	82	3	3	12	27	87	82
United Kingdom	19	6	4	16	4	21	23
Germany	10	34	20	1	14	5	10
Czech Republic	64	38	61	2	31	78	76
Slovakia	58	37	60	3	24	56	57
Malaysia	8	10	6	20	2	8	6
Austria	39	64	54	64	64	1	51



Таблица 2.13 – Рейтинг стран, экспортирующих газ в газообразном состоянии, по индексам центральности на 2018 год

Country	InDeg	OutDeg	Deg	Clos	BetwCentr	PageRank	EigenVec
Germany	1	20	7	27	2	8	1
China	2	12	6	1	12	17	27
Ukraine	3	35	16	35	35	9	3
United Kingdom	4	7	5	31	8	11	8
Kazakhstan	40	1	1	10	13	46	40
Hungary	33	2	2	36	5	15	33
Belgium	11	3	3	28	3	25	10
Uzbekistan	43	4	4	22	20	49	43
Azerbaijan	42	10	12	2	15	48	42
United States of America	12	13	14	3	16	1	13
Area Nes	7	8	10	12	1	4	2
Czech Republic	21	6	9	11	4	16	6
Lithuania	14	21	20	14	19	2	16

Таблица 2.14 – Рейтинг стран, экспортирующих СПГ, по индексам центральности на 2018 год

Country	InDeg	OutDeg	Deg	Clos	BetwCentr	PageRank	EigenVec
China	1	31	9	38	18	2	10
Korea, Republic of	2	32	10	40	20	5	1
Japan	3	33	13	48	28	14	11
France	4	34	17	41	21	6	2
United Kingdom	5	15	8	20	2	15	3
United States of America	22	1	1	10	1	25	20
Norway	31	2	2	15	3	18	26
Nigeria	40	3	3	16	10	41	47
India	10	20	21	1	14	21	22
Area Nes	25	18	25	3	12	1	37
Portugal	11	37	26	39	19	3	5
Sweden	24	17	24	25	8	4	23
Pakistan	7	36	19	45	25	10	4

В динамике индексы центральности для определенной страны очень изменчивы, так как на протяжении практически двадцатилетнего рассматриваемого периода страны – экспортеры меняли покупателей сырья, а страны – импортеры переключались на других поставщиков сырья. Причиной этому выступали экономические и политические изменения, а также истощение ресурсов на разведанных местах добычи и нехватка финансирования на разработку новых месторождений, изменение потребления внутри стран – некоторые страны были вынуждены прекратить экспорт, так как спрос внутри страны увеличился.

## 2.4 Оценивание схожести слоёв сети торговли продовольствием

### 2.4.1 Описание модели схожести слоёв сети

В данном разделе представим модель по оценки схожести слоёв в сетевых структурах, основанную на сравнении связей каждой из вершин.

Вершинами  $N = \{1, \dots, n\}$  в рассматриваемой сети являются страны, а рёбрами – взвешенные направленные потоки экспорта/импорта продовольствия в тыс. долларах. Ввиду того, что объёмы потоков по различным типам продовольствия  $S = \{1, \dots, s\}$  сильно отличаются друг от друга, необходимо провести нормализацию сети. Данный шаг необходим для того, чтобы различные слои можно было сравнивать между собой.

Существует множество различных подходов для нормализации данных и приведении их к одной шкале. В данной работе использовалась нормализация, описанная в работах [83], [96], которая учитывает индивидуальные атрибуты вершин и возможность группового влияния. Каждой вершине на каждом слое ставится в соответствие пороговое значение, обозначающее критическую величину потери входящего суммарного импорта. В нашей работе данный порог вычисляется по формуле

$$threshold_i^j = q\% \cdot \max(import_i^j, export_i^j)$$

для каждой вершины  $i = 1, \dots, n$  на каждом слое  $j = 1, \dots, s$ .

После вычисления пороговых значений для каждой вершины вычисляется сила прямого влияния вершин друг на друга с учётом групповых взаимодействий согласно [83, 96]. Таким образом, взвешенные рёбра на каждом слое принимают значения от 0 до 1, что позволяет нам сравнивать различные слои в сети торговли продовольствием. Также стоит отметить, что часть слабых рёбер с помощью приведённой нормировки убираются из рассмотрения, что делает каждый слой более разреженным, а последующий анализ – вычислительно проще.

Теперь рассмотрим подход, основанный на сравнении векторов связей каждой вершины. Логично предположить, что два слоя будут считаться наиболее схожими, если одна и та же вершина в каждом слое связана с другими вершинами наиболее похожим образом. В данном подходе особенно важно учитывать направления и силу связи между вершинами, которую мы получили после шага нормализации сети.

Формализуем данный подход. Рассмотрим два слоя сети  $G_1 \in S$  и  $G_2 \in S$ , представляющих собой отдельные графовые структуры. Каждому слою соответствует своя матрица смежности  $A(G_1)$  и  $A(G_2)$ , показывающие силу связи  $a_{ij}(G_k)$  между вершинами  $i \in N$  и  $j \in N$  в графе  $G_k, k \in \{1,2\}$ . Обозначим за вектор  $A_{i*}(G_k)$   $i$ -ую строку соответствующей матрицы смежности, из которой предварительно уберём  $i$ -ый столбец. Данный вектор показывает, как страна  $i$  влияет на остальные страны в рассматриваемой сети. Теперь мы можем посчитать корреляцию двух векторов для одной и той же вершины в различных слоях, как  $corr(A_{i*}(G_1), A_{i*}(G_2))$ . После чего меру схожести двух слоёв можно оценить как нормированную сумму корреляций исходящих связей по всем вершинам, т.е.

$$SIM_{outdeg}(G_1, G_2) = \frac{\sum_{i=1}^n corr(A_{i*}(G_1), A_{i*}(G_2))}{n}.$$

В такой форме данный подход имеет ряд недостатков. Например, равнозначно учитывается влияние слабых элементов сетевой структуры, чьи значения связей бывают очень нестабильными от слоя к слою, что может сильно

искажить результат. Поэтому, выберем только важные вершины, согласно некоторому показателю, полученному либо из самой сетевой структуры, либо из внешних источников данных. В данной работе было выбрано 40 стран с наибольшим показателем ВВП за каждый отдельный год. Таким образом, вместо того, чтобы сравнивать все вершины в сетевой структуре, мы выбрали только подграф нормированной сети из ключевых участников и связей между ними. Схематично данный подход показан на рисунке 2.7.

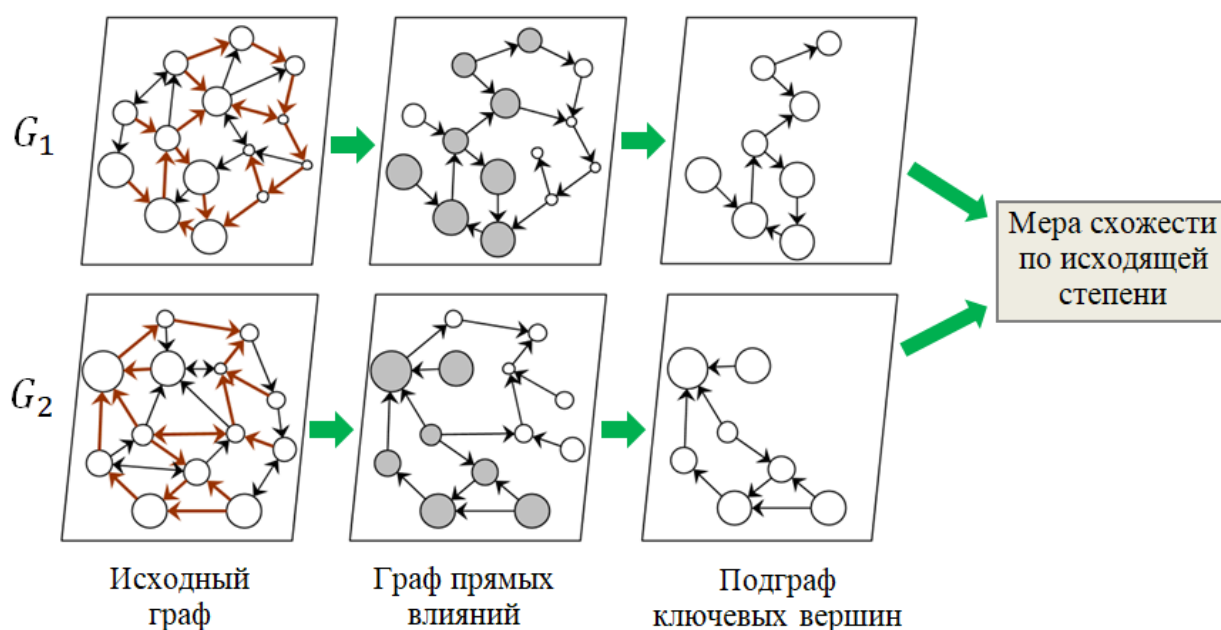


Рисунок 2.7 – схема сравнения слоёв графа

Отметим, что данный вид сравнения слоёв сети применим как к различным подграфам экспорта/импорта продовольствия, так и к различным временным срезам конкретного вида товара.

Более подробно приведённая в отчёте модель описана в статье [119].

#### 2.4.2 Применение модели к сети торговли продовольствием

Основным источником данных по международной торговле является база данных World Integrated Trade Solution (WITS Comtrade) [120]. WITS Comtrade предоставляет двустороннюю статистику по международной торговле по различным категориям товаров и услуг. В данной работе было выбрано 33 продуктовых категории согласно классификации SITC за период с

1996 по 2017 годы. В результате мы получили сетевую структуру, содержащую 726 слоёв, где каждый слой соответствует определённой категории товара за конкретный год.

В Таблице 2.15 приведены результаты применения описанной модели сравнения слоёв сети к различным категориям товаров за 1996, 2002, 2008, 2013 и 2017 годы. За каждый рассматриваемый период было выбрано 5 наиболее схожих пар слоёв.

Таблица 2.15 – 5 наиболее схожих слоёв сети торговли продовольствием

1996	2002	2008	2013	2017
Fresh fruit vs. Fruit prep (0.542)	Cereal prep vs. Food prep (0.620)	Cereal prep vs. Chocolate (0.645)	Cereal prep vs. Chocolate (0.557)	Cereal prep vs. Food prep (0.537)
Cereal prep vs. Food prep (0.512)	Cereal prep vs. Chocolate (0.620)	Cereal prep vs. Food prep (0.579)	Fresh fruit vs. Fresh veg (0.554)	Cereal prep vs. Chocolate (0.522)
Fruit prep vs. Veg prep (0.507)	Fruit prep vs. Veg prep (0.522)	Fresh veg vs. Sugar prep (0.524)	Cereal prep vs. Food prep (0.505)	Fresh fruit vs. Fruit prep (0.506)
Cereal prep vs. Sugar prep (0.433)	Chocolate vs. Food prep (0.493)	Fruit prep vs. Sugar prep (0.517)	Fresh fruit vs. Fruit prep (0.491)	Sugar prep vs. Chocolate (0.499)
Fresh fish vs. Fresh fruit (0.429)	Fresh veg vs. Sugar prep (0.473)	Chocolate vs. Food prep (0.515)	Fruit prep vs. Fresh veg (0.463)	Fresh fruit vs. Fresh veg (0.495)

Примечание - \*Названия соответствуют официальной кодировке SITC

Как видно из Таблицы 2.15 наиболее схожими являются группы однородных товаров (фрукты и овощи, шоколад и сахар, и т.д.). Основной причиной такой высокой схожести данных групп товаров является то, что эти пары товаров производятся в схожей среде и, если у страны есть возможность экспортировать один товар, следовательно, у этой страны также имеется возможность производить и экспортировать схожий по категории товар. Однако можно заметить, что разнородные группы товаров также могут иметь высокую схожесть согласно метрике влияния стран. Например, такие продукты, как злаковые и шоколад являются очень близкими с точки зрения метрики влияния стран, так как согласно данным ФАО [121] их основными экспортёрами являются такие сильные экономики, как Германия, Бельгия и США.

## 2.5 Исследование многосторонних торговых потоков в гравитационной модели торговли

В данном подразделе отчета приведены результаты применения модификации гравитационной модели к анализу многосторонних торговых потоков. Модификация состоит во включении в гравитационную модель индексов центральности сети. Чаще всего индексы центральности добавляют в модель как признаки в регрессии. В нашем исследовании предлагается брать также и отношения индексов центральности стран, как экспортеров, так и импортеров. Такой подход позволит лучше оценить важность данной страны для мировой торговли в сравнении с ее торговыми партнерами. Более того, рассматривая разницу между центральностями торговых партнеров можно получить больше информации о возможных двусторонних барьерах между ними.

Для оценки полученной модели использовалась база данных ВАСИ, представленная в [122]. Данные по экспорту и импорту были сведены к единому значению потоков между двумя странами. Очень часто наблюдается явление несоответствия друг другу экспортного потока из страны  $i$  в страну  $j$  и импорта в  $j$  из  $i$ , хотя они должны совпадать, если их очистить от тарифных и транспортных издержек. Методика сведения таких потоков к единой величине подробно изложена в работе [122]. В ее основе лежит использование взвешенной суммы импортных и экспортных потоков на основании надежности страны с точки зрения того, насколько достоверные данные по объемам торговли она в среднем предоставляет.

В рамках данного исследования рассматривалась статичная модель международной торговли за 2018 год. Для более точного описания каждого торгового потока импортные и экспортные потоки не агрегируются, а рассматриваются по отдельности. Гравитационная модель оценивается по 19024 торговым потокам, существующим между 177 суверенными и признанными государствами. Всего в рамках международной торговой сети в данном году существовало 31152 торговых потоков. В Таблице 2.16 представлены дескриптивные статистики распределения объема торговых потоков.

Таблица 2.16 – Дескриптивные статистики торговых потоков за 2018 год

Среднее	Стан. Откл.	Минимальное значение	25%	50%	75%	Максимальное значение
893 млн.	7562 млн.	1000.0	355402	7,68 млн.	108 млн.	504586 млн.

Как видно из Таблицы 2.16 данные по потокам крайне неоднородные, однако взятие логарифма существенно снижает дисперсию в данных.

В качестве индексов центральности предлагается использовать коэффициенты центральности [123]. К примеру, центральность данной вершины по входящим в нее ребрам в рамках экономической терминологии можно рассматривать, как центральность страны как импортера, а центральность по исходящим ребрам как центральность страны в качестве экспортера. Последняя показывает конкурентоспособность страны на мировом рынке – чем выше ее центральность как экспортера, тем больше она экспортирует своей продукции в другие страны. С другой стороны, центральность страны как импортера показывает покупательную способность потребителей – чем выше эта центральность, тем больше товаров из-за границы готовы приобрести потребители данной страны. Рассматривая центральности с данной точки зрения, не имеет смысла добавлять в модель сами центральности напрямую. Вместо этого, в случае если оценивается торговый поток из страны  $i$  в страну  $j$ , можно добавить отношение экспортного потенциала страны  $i$  к импортному потенциалу страны  $j$ . Так как чем выше экспортный потенциал, тем выше качество производимых в стране  $i$  товаров. Высокий же импортный потенциал указывает на то, что в стране  $j$  потребители готовы приобрести более качественные товары. Если отношение двух центральностей близко к 1, то значит, наблюдается случай, когда импортирующая страна приобретает наиболее предпочтительные ею товары. Иначе же, если отношение центральностей сильно отклоняется от 1, существуют барьеры в торговле между двумя странами на уровне потребителей – они не будут предъявлять спрос либо на слишком высококачественные

товары (если отношение сильно выше 1), либо на слишком низкокачественные (если отношение сильно меньше 1).

Далее проводилась эконометрическая оценка модели по следующим признакам: ВВП страны-экспортера, ВВП страны-импортера, расстояние между странами и отношение центральных. Результаты оценки приведены в Таблице 2.17.

Таблица 2.17 – Эконометрическая оценка полученного уравнения.

	Точечная оценка	Стан. Откл.	t-статистика	p-value
Константа	-10.1636	0.147	-68.907	0.000
Расстояние	-1.2691	0.024	-53.366	0.000
ВВП экс.	1.2257	0.009	142.209	0.000
ВВП имп.	0.9354	0.008	110.200	0.000
Отношение ин-дексов	-0.5568	0.087	-6.393	0.000

Для сравнения, в Таблице 2.18 приведена оценка уравнения без отношения индексов центральных:

Таблица 2.18 – Оценка базового уравнения.

	Точечная оценка	Стан. Откл.	t-статистика	p-value
Константа	-28.2285	0.390	-72.330	0.000
Расстояние	-1.2963	0.023	-55.326	0.000
ВВП экс.	1.2382	0.008	147.305	0.000
ВВП имп.	0.9483	0.008	114.897	0.000

Коэффициент детерминации в обоих случаях составляет 0,69. Также важно отметить тот факт, что коэффициент при константе в расширенной модели существенно ниже. Большая отрицательная величина в базовой модели приводила к тому, что оценки объемов торговли между маленькими странами получались отрицательными. В расширенной модели значение коэффициента меньше более чем в 2,5 раза.



## **2.6 Оценивание центральностей многослойной сети методами теории функций доверия**

В этом подразделе отчета приведены результаты по оцениванию центральностей многослойной сети методами теории функций доверия.

### **2.6.1 Постановка задачи**

Будем рассматривать многослойные сети, ребра которых имеют векторные веса. Требуется оценить центральности вершин такой сети. Теоретически имеется несколько возможностей это сделать. Первая возможность состоит в том, что нужно найти центральности вершин для каждого слоя в отдельности, а затем агрегировать эти центральности. Вторая возможность – локально агрегировать атрибуты (например, веса рёбер) слоев в скалярные величины, а затем оценить центральности полученного нового взвешенного графа.

Ниже приведены предварительные результаты другого подхода, отличного от вышеописанных. А именно, каждому атрибуту и каждой вершине ставится в соответствие функция множеств, с помощью которой оценивается центральность графа. В случае многослойных сетей мы имеем набор таких функций множеств для каждой вершины. Эти функции могут быть агрегированы в одну функцию, после чего вычисляются центральности вершин с помощью новых функций множеств.

Развитый инструментарий комбинирования функций множеств имеет теория функций доверия (теории свидетельств Демпстера-Шейфера [124], [125]).

Методы теории функций доверия применяются при решении различных прикладных задач. В последние годы эта теория находит применение и в сетевом анализе данных. Так, в [126] в качестве тела свидетельств фиксированной вершины графа рассматриваются три фокальных элемента, соответствующих уровню состояния центральности вершины – высокий, низкий, высокий или низкий. В этой работе тела свидетельств степенных и весовых центральностей агрегировались с помощью правила комбинирования Демпстера. В [127] пра-

вило Демпстера использовалось для агрегирования индикаторов влияния в социальной сети. В [128] предложен алгоритм выделения сообществ в сетях, основанный на введении мер локальной плотности графа в фиксированной вершине и детектировании пиков таких мер (density peaks).

Особенностью нашего подхода является то, что агрегируются функции множеств, задаваемые на множестве вершин графа, т.е. на первичных данных, а не на множестве уровней состояния, индикаторов и т.д., т.е. на вторичных данных, как в цитированных выше работах.

Заметим, что исследование функций множеств на множестве вершин в сетевом анализе восходит к работам Myerson [79].

### 2.6.2 Основные сведения из теории функций доверия

Пусть  $X$  – некоторое конечное множество, которое называют фреймом различения,  $2^X$  – множество всех его подмножеств. В теории функций доверия (теории свидетельств Демпстера-Шейфера [124], [125]) рассматривается функция множеств  $m: 2^X \rightarrow [0,1]$ , удовлетворяющая условиям нормировки:  $m(\emptyset) = 0$  и  $\sum_{A \in 2^X} m(A) = 1$ , которая называется базовым вероятностным назначением или функцией масс. Те подмножества  $A \in 2^X$ , для которых  $m(A) > 0$  называются фокальными элементами, а совокупность всех фокальных элементов вместе с их функциями масс – телом свидетельства.

С функцией масс  $m$  можно связать две функции множеств – функцию доверия  $g(A) = \sum_{B \subseteq A} m(B)$  и (двойственную ей) функцию правдоподобия  $\bar{g}(A) = 1 - g(\neg A) = \sum_{B: A \cap B \neq \emptyset} m(B)$ . Эти функции можно рассматривать как нижнюю и верхнюю оценки вероятности события  $A$ :  $g(A) \leq P(A) \leq \bar{g}(A)$ ,  $A \in 2^X$ . Значение функции доверия  $g(A)$  можно трактовать как степень доверия к тому, что истинная альтернатива  $x \in A$ , если известно, что  $x \in X$ . Пусть  $Bel(X)$  множество всех функций доверия на  $X$ .

Функцию доверия  $g$  удобно представлять с помощью выпуклой комбинации (с коэффициентами  $\{m(A)\}_A$ ) так называемых категориальных функций

доверия:  $\eta_B(A) = \begin{cases} 1, & B \subseteq A, \\ 0, & B \not\subseteq A, \end{cases} \quad B \in 2^X \setminus \{\emptyset\}$ . Тогда  $g = \sum_B m(B)\eta_B$ . Заметим, что

$\eta_X$  описывает бессодержательное свидетельство, что  $x \in X$ . Поэтому такая функция доверия называется бессодержательной.

Зная функцию доверия  $g$  можно восстановить функцию масс с помощью преобразования Мёбиуса  $m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} g(B)$ .

В теории функций доверия хорошо развит инструментальный агрегирования (комбинирования). Под правилом комбинирования двух функций доверия  $g_1, g_2 \in \text{Bel}(X)$  понимают некоторый оператор  $R: (g_1, g_2) \mapsto g$ ,  $g, g_1, g_2 \in \text{Bel}(X)$ . Комбинирование соответствующих тел свидетельств будем обозначать через  $m = m_1 \mathcal{D}_R m_2$ . В теории и приложениях рассматривают различные правила комбинирования. Исторически первым и наиболее популярным в приложениях из таких правил является правило Демпстера [124]:

$$m(A) = (m_1 \mathcal{D}_D m_2)(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_1(B)m_2(C) \text{ для всех } A \neq \emptyset \text{ и } m(\emptyset) = 0,$$

где  $K = \sum_{B \cap C = \emptyset} m_1(B)m_2(C)$ . Величина  $K = K(m_1, m_2) \in [0, 1]$  характеризует величину конфликта между свидетельствами. Если  $K=1$  (абсолютный конфликт), то правило Демпстера неприменимо.

Близким к правилу Демпстера является правилу Ягера [129]:

$$m(A) = (m_1 \mathcal{D}_Y m_2)(A) = \sum_{B \cap C = A} m_1(B)m_2(C), \quad A \neq \emptyset, X, \quad m(\emptyset) = 0,$$

$$m(X) = K + m_1(X)m_2(X).$$

Другими словами, в правиле Ягера величина конфликта  $K$  перераспределяется в массу «незнания»  $m(X)$ .

Обзор других правил комбинирования можно найти в [130].

Если различные источники информации, описываемые с помощью функций доверия, имеют разные надежности или различный приоритет, то это можно учесть с помощью метода дисконтирования, который был предложен в [125]. Дисконтирование функции масс  $m$  осуществляется с помощью коэффициента  $\alpha \in [0,1]$  по следующему правилу:

$$m^\alpha(A) = (1 - \alpha)m(A) \text{ для } A \neq X \text{ и } m^\alpha(X) = (1 - \alpha)m(X) + \alpha.$$

Если  $\alpha = 0$ , то источник информации считается абсолютно надежным. В этом случае  $m^\alpha(A) = m(A)$  для всех  $A \in 2^X$ . Если же  $\alpha = 1$ , то  $m^\alpha(X) = 1$  (т.е. соответствующая функция доверия будет бессодержательной:  $g = \eta_X$ ).

### 2.6.3 Основные обозначения и классические центральности графа

Ниже будем рассматривать неориентированный связный взвешенный граф  $G = (V, E, W)$ , где  $V = \{v_k\}$  – множество вершин графа,  $E = \{e_{ij}\}$  – множество ребер графа ( $e_{ij} = e(v_i, v_j)$  – ребро графа между вершинами  $v_i$  и  $v_j$ ),  $W = (w_{ij})$  – матрица весов графа,  $w_{ij} = w(v_i, v_j)$  – вес ребра графа между вершинами  $v_i$  и  $v_j$ . Будем предполагать, что все веса графа  $w_{ij} \in [0,1]$ , причем  $w_{ij} = 0 \Leftrightarrow e_{ij} \notin E$ . Числа  $w_{ij}$  можно интерпретировать, как степень взаимодействия между вершинами  $i$  и  $j$ . Будем считать, что  $w_{ii} = 1$  для всех индексов  $i$ .

Существует несколько подходов к оценке влияния (центральности, активности и т.д.) вершин в сетях (графах) [131]. Например, простейшей степенной центральностью является степень  $j$ -й вершины  $n_j$  графа, т.е. количество ребер, инцидентных  $j$ -й вершине. Обобщением простейшей степенной центральности является так называемая собственная векторная центральность (eigenvector centrality), которая учитывает не только ближние, но и дальние связи вершины. Вычисление такой меры центральности для каждой вершины связано с решением задачи на собственные значения относительно матрицы смежности  $A$  (или матрицы весов  $W$  в случае взвешенного графа) графа сети: вектор отно-

сительных центральностей  $\mathbf{x}$  является собственным вектором матрицы смежности, соответствующему наибольшему собственному значению  $\lambda_{\max}$ , т.е.  $W\mathbf{x} = \lambda_{\max}\mathbf{x}$ . Такой способ вычисления центральности был предложен в [132].

Пример 2.1. Для графа, представленного на рисунке 2.8,

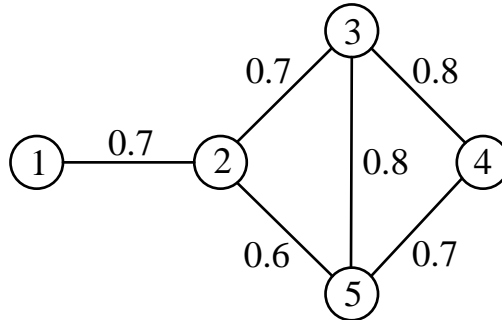


Рисунок 2.8 – Пример взвешенного графа

его матрица весов  $W$  равна

$$W = \begin{pmatrix} 0 & 0.7 & 0 & 0 & 0 \\ 0.7 & 0 & 0.7 & 0 & 0.6 \\ 0 & 0.7 & 0 & 0.8 & 0.8 \\ 0 & 0 & 0.8 & 0 & 0.7 \\ 0 & 0.6 & 0.8 & 0.7 & 0 \end{pmatrix}.$$

Тогда, вычисляя собственную векторную центральность, мы получим, что  $\lambda \approx 1.9$  и соответствующий собственный вектор с положительными координатами равен  $\mathbf{x} = (0.16; 0.43; 0.56; 0.43; 0.53)$ . Т.е. центральности вершин графа упорядочены следующим образом:  $x_1 \leq x_2 = x_4 \leq x_5 \leq x_3$ .

#### 2.6.4 Оценивание центральностей однослойной сети методами теории функций доверия

Центральность вершин графа зависит от относительного суммарного «веса» смежных с данной вершиной подграфов. А суммарный вес отдельных подграфов можно описать с помощью монотонных мер, в том числе, с помощью мер доверия.

Рассмотрим на множестве всех вершин графа  $V$  семейство функций доверия  $g_k = \sum_B m_k(B) \eta_B$ ,  $k = 1, \dots, |V|$ , где функция  $g_k$  связана с  $k$ -й вершиной

графа. Пусть  $N_k^{(p)} = N^{(p)}(v_k)$   $p$  – окрестность вершины  $v_k$ , т.е. множество вершин графа, находящихся на расстоянии не больше  $p$  от вершины  $v_k$  (считаем, что  $v_k \notin N^{(p)}(v_k)$ ),  $|W| = \sum_{i < j} w_{ij}$  – сумма всех весов графа (вес графа). Определим функцию масс  $m_k$  следующим образом:

$$m_k(A) = \frac{1}{|W|} \begin{cases} w(u, v_k), & \text{если } A = \{u\} \subseteq N_k^{(1)}, \\ \gamma_k(u, v)w(u, v), & \text{если } A = \{u, v\} \subseteq N_k^{(p)} \text{ \& } e(u, v) \in E, \\ |W| - \sum_{u \in N_k^{(1)}} w(u, v_k) - \sum_{\substack{u, v \in N_k^{(p)}, \\ e(u, v) \in E}} \gamma_k(u, v)w(u, v), & \text{если } A = V, \\ 0 & \text{в остальных случаях,} \end{cases} \quad (2.1)$$

где  $\gamma_k(u, v) \in [0, 1]$  – коэффициент дисконтирования, который понижает значимость дальних от вершины  $v_k$  ребер. Например, можно взять

$$\gamma_k(u, v) = \frac{1}{1 + \min\{d(u, v_k), d(v, v_k)\}}, \quad (2.2)$$

где  $d$  – расстояние между вершинами.

Нетрудно видеть, что по построению функции множеств  $m_k : 2^V \rightarrow [0, 1]$  удовлетворяет условиям нормировки:  $m_k(\emptyset) = 0$ ,  $\sum_{A \in 2^V} m_k(A) = 1$ . Поэтому  $m_k$  – базовые вероятностные назначения (функции масс), а  $g_k = \sum_B m_k(B) \eta_B$  – функции доверия на  $V$ .

Ближкие к выше введенным функциям множеств конструкции для оценивания влияния вершин графа вводились в [133].

Тело свидетельств  $\{m_k(B)\}$  характеризует распределение «чистого» взаимодействия вершины  $v_k$  с подмножеством близких к  $v_k$  вершин множеств  $B$ , т.е. взаимодействия за вычетом взаимодействия с другими подмножествами

множества  $B$ . Величина  $m_k(V)$  – это величина «незнания» относительно взаимодействия  $k$ -й вершины с дальними вершинами и их подмножествами, находящиеся за пределами окрестности  $N_k^{(p)}$ .

Взаимодействия вершины  $v_k$  с другими вершинами графа можно оценить, если перераспределить значения масс несинглетонов равномерно по синглетонам. Это можно сделать с помощью, так называемых пигнистических вероятностей (pignistic probability), введенных в [134]:

$$Bet_{v_k}(\{v_i\}) = Bet_k(\{v_i\}) = \sum_{A: v_i \in A} \frac{m_k(A)}{|A|}.$$

Известно, что пигнистическая вероятность для функции доверия  $g_k = \sum_B m_k(B)\eta_B$ , определенной на  $V$ , совпадает со значениями Шепли [135], которые широко используются в теории коалиционных (кооперативных) игр и вычисляются по формуле

$$Bet_k(\{v_i\}) = \sum_{A \subseteq V, v_i \in A} \frac{(|V| - |A|)!(|A| - 1)!}{|V|!} (g_k(A) - g_k(A \setminus \{v_i\})), \quad i=1, \dots, |V|.$$

Величина  $Bet_k(\{v_i\})$  характеризует долю активности  $k$ -й вершины, приходящуюся на  $i$ -ю вершину. Тогда величина

$$q(v_i) = q_i = \sum_{u \in V} Bet_u(\{v_i\}), \quad i=1, \dots, |V| \quad (2.3)$$

будет характеризовать суммарное взаимодействие с  $i$ -й вершиной.

Пример 2.2. Для графа рисунка 2.8 сумма весов равна  $|W| = 4.3$ . Найдем функции доверия вершин этого графа для 2-окрестности по формулам (2.1) и (2.2). Имеем

$$g_1 = \frac{7}{43}\eta_{\{v_2\}} + \frac{7}{86}\eta_{\{v_2, v_3\}} + \frac{3}{43}\eta_{\{v_2, v_5\}} + \frac{8}{129}\eta_{\{v_3, v_5\}} + \frac{161}{258}\eta_V,$$

$$g_2 = \frac{7}{43}\eta_{\{v_1\}} + \frac{7}{43}\eta_{\{v_3\}} + \frac{6}{43}\eta_{\{v_5\}} + \frac{4}{43}\eta_{\{v_3, v_4\}} + \frac{4}{43}\eta_{\{v_3, v_5\}} + \frac{7}{86}\eta_{\{v_4, v_5\}} + \frac{23}{86}\eta_V,$$

$$g_3 = \frac{7}{43}\eta_{\{v_2\}} + \frac{8}{43}\eta_{\{v_4\}} + \frac{8}{43}\eta_{\{v_5\}} + \frac{7}{86}\eta_{\{v_1, v_2\}} + \frac{3}{43}\eta_{\{v_2, v_5\}} + \frac{7}{86}\eta_{\{v_4, v_5\}} + \frac{10}{43}\eta_V,$$

$$g_4 = \frac{8}{43}\eta_{\{v_3\}} + \frac{7}{43}\eta_{\{v_5\}} + \frac{7}{86}\eta_{\{v_2, v_3\}} + \frac{3}{43}\eta_{\{v_2, v_5\}} + \frac{4}{43}\eta_{\{v_3, v_5\}} + \frac{35}{86}\eta_V,$$

$$g_5 = \frac{6}{43}\eta_{\{v_2\}} + \frac{8}{43}\eta_{\{v_3\}} + \frac{7}{43}\eta_{\{v_4\}} + \frac{7}{86}\eta_{\{v_1, v_2\}} + \frac{7}{86}\eta_{\{v_2, v_3\}} + \frac{4}{43}\eta_{\{v_3, v_4\}} + \frac{11}{43}\eta_V.$$

Тогда матрица пигнистических вероятностей  $Bet = (Bet_k(\{v_i\}))$  (в  $k$ -й строке вероятности меры  $Bet_k$ ) будет равна

$$Bet = \begin{pmatrix} 0.125 & 0.363 & 0.197 & 0.125 & 0.19 \\ 0.216 & 0.053 & 0.309 & 0.141 & 0.28 \\ 0.087 & 0.285 & 0.047 & 0.273 & 0.308 \\ 0.081 & 0.157 & 0.355 & 0.081 & 0.326 \\ 0.092 & 0.272 & 0.324 & 0.26 & 0.051 \end{pmatrix}.$$

Вектор индексов взаимодействия  $\mathbf{q} = (q_1, \dots, q_5)$  будет равен  $(0.602, 1.13, 1.232, 0.881, 1.155)$ . Таким образом, центральности вершин графа будут упорядочены следующим образом:  $x_1 \leq x_4 \leq x_2 \leq x_5 \leq x_3$ .

Можно доказать следующее утверждение о связи суммарного взаимодействия, вычисленного по формуле (2.3) с классическими центральностями в простейшем случае, когда центральности вычисляются с помощью 1-окрестности, причем порожденный окрестностью  $N^{(1)}$  подграф является звездой.

**Теорема 2.1.** *Если суммарное взаимодействие в графе  $G=(V, E, W)$  вычисляется с помощью 1-окрестности и подграф, порожденный окрестностью  $N^{(1)}(v)$ , является звездой, то суммарное взаимодействие с вершиной  $v$  будет равно*

$$q(v) = \frac{n(v)}{|W|} + 1 - \frac{2}{|V|},$$

где  $n(v) = \sum_{u: e(v,u) \in E} w(v,u)$  – сумма весов всех ребер, инцидентных вершине  $v$ .



Из этого утверждения вытекает, что в указанном случае суммарное взаимодействие с точностью до постоянного слагаемого, независящего от вершины, пропорционально весовой степени  $n(v)$ .

### 2.6.5 Многослойные сети и агрегирование центральностей

Рассмотрим теперь граф многослойной сети  $G=(V,E,W)$ , где  $V=\{v_k\}$  – множество вершин графа,  $E=\{e_{ij}\}$ , – множество ребер графа ( $e_{ij}=e(v_i,v_j)$  – ребро графа между вершинами  $v_i$  и  $v_j$ ),  $W=(w_{ij})$  – матрица весов многослойной сети,  $w_{ij}=(w_{ij}^{(1)},...,w_{ij}^{(l)})$ ,  $w_{ij}^{(s)}=w^{(s)}(v_i,v_j)$  – вес ребра  $s$ -го слоя сети между вершинами  $v_i$  и  $v_j$ .

Пусть  $W^{(s)}=(w_{ij}^{(s)})$  матрица весов  $s$ -го слоя сети  $G$ ,  $s=1,...,l$ . Тогда можно описанным выше способом найти центральности всех вершин графа  $G=(V,E,W^{(s)})$ ,  $s=1,...,l$ . В результате получим векторные центральности вершин  $\mathbf{q}_i=(q_i^{(1)},...,q_i^{(l)})$ ,  $i=1,...,|V|$ . Тогда для определения центральных вершин возникнет задача упорядочивания векторов  $\mathbf{q}_i$ . Вместо этого рассмотрим другой способ.

Для каждого графа  $G=(V,E,W^{(s)})$ ,  $s=1,...,l$  найдем соответствующий ему набор тел свидетельств  $\{m_k^{(s)}\}_{k=1}^{|V|}$ . Далее с помощью правила комбинирования свидетельств  $R$  и (возможно) коэффициентов дисконтирования  $\alpha=(\alpha_1,...,\alpha_l)$ , учитывающих приоритеты слоёв при оценивании центральностей, найдем агрегированное свидетельство  $m_k=\bigwedge_{s=1}^l \alpha m_k^{(s)}$ . После чего для этого агрегированного свидетельства описанным в предыдущем пункте способом найдем центральности всех вершин графа.

Пример 2.3. Предположим, что граф из примера 2.1 имеет два слоя, матрицы весов которых равны

$$W^{(1)} = \begin{pmatrix} 0 & 0.7 & 0 & 0 & 0 \\ 0.7 & 0 & 0.7 & 0 & 0.6 \\ 0 & 0.7 & 0 & 0.8 & 0.8 \\ 0 & 0 & 0.8 & 0 & 0.7 \\ 0 & 0.6 & 0.8 & 0.7 & 0 \end{pmatrix}, \quad W^{(2)} = \begin{pmatrix} 0 & 0.8 & 0 & 0 & 0 \\ 0.8 & 0 & 0.5 & 0 & 0.9 \\ 0 & 0.5 & 0 & 0.7 & 0.8 \\ 0 & 0 & 0.7 & 0 & 0.6 \\ 0 & 0.9 & 0.8 & 0.6 & 0 \end{pmatrix}$$

(матрица  $W^{(1)}$  совпадает с матрицей  $W$  из предыдущего примера). Тогда набор функций доверия, соответствующий графу  $G = (V, E, W^{(1)})$  получен в предыдущем примере. Обозначим его через  $\{g_k^{(1)}\}_{k=1}^5$ .

Найдем теперь набор функций доверия, соответствующий графу  $G = (V, E, W^{(2)})$ . Сумма весов равна  $|W^{(2)}| = 4.3$ . Тогда

$$g_1^{(2)} = \frac{8}{43}\eta_{\{v_2\}} + \frac{5}{86}\eta_{\{v_2, v_3\}} + \frac{9}{86}\eta_{\{v_2, v_5\}} + \frac{8}{129}\eta_{\{v_3, v_5\}} + \frac{76}{129}\eta_V,$$

$$g_2^{(2)} = \frac{8}{43}\eta_{\{v_1\}} + \frac{5}{43}\eta_{\{v_3\}} + \frac{9}{43}\eta_{\{v_5\}} + \frac{7}{86}\eta_{\{v_3, v_4\}} + \frac{4}{43}\eta_{\{v_3, v_5\}} + \frac{3}{43}\eta_{\{v_4, v_5\}} + \frac{21}{86}\eta_V,$$

$$g_3^{(2)} = \frac{5}{43}\eta_{\{v_2\}} + \frac{7}{43}\eta_{\{v_4\}} + \frac{8}{43}\eta_{\{v_5\}} + \frac{4}{43}\eta_{\{v_1, v_2\}} + \frac{9}{86}\eta_{\{v_2, v_5\}} + \frac{3}{43}\eta_{\{v_4, v_5\}} + \frac{23}{86}\eta_V,$$

$$g_4^{(2)} = \frac{7}{43}\eta_{\{v_3\}} + \frac{6}{43}\eta_{\{v_5\}} + \frac{5}{86}\eta_{\{v_2, v_3\}} + \frac{9}{86}\eta_{\{v_2, v_5\}} + \frac{4}{43}\eta_{\{v_3, v_5\}} + \frac{19}{43}\eta_V,$$

$$g_5^{(2)} = \frac{9}{43}\eta_{\{v_2\}} + \frac{8}{43}\eta_{\{v_3\}} + \frac{6}{43}\eta_{\{v_4\}} + \frac{4}{43}\eta_{\{v_1, v_2\}} + \frac{5}{86}\eta_{\{v_2, v_3\}} + \frac{7}{86}\eta_{\{v_3, v_4\}} + \frac{10}{43}\eta_V.$$

Матрица пигнистических вероятностей  $Bet_k^{(2)}(\{v_i\})$  этого набора функций доверия, соответствующих вторым атрибутам, будет равна

$$Bet^{(2)} = \begin{pmatrix} 0.118 & 0.385 & 0.178 & 0.118 & 0.201 \\ 0.235 & 0.049 & 0.252 & 0.124 & 0.34 \\ 0.1 & 0.269 & 0.053 & 0.251 & 0.327 \\ 0.088 & 0.17 & 0.327 & 0.088 & 0.327 \\ 0.093 & 0.331 & 0.302 & 0.227 & 0.047 \end{pmatrix}.$$

Вектор индексов влияния  $\mathbf{q}^{(2)}$  для вторых атрибутов будет равен (0.634, 1.204, 1.113, 0.809, 1.241). Таким образом, центральности вершин относительно второго атрибута будут упорядочены следующим образом:  $x_1 \leq x_4 \leq x_3 \leq x_2 \leq x_5$ .

Теперь найдем попарную агрегацию функций доверия  $g_k^{(1)}$  и  $g_k^{(2)}$  по правилу Демпстера  $D$ . В результате получим новые функции доверия

$g_k = D(g_k^{(1)}, g_k^{(2)})$ , характеризующие распределение взаимодействий вершин двухатрибутного графа. Эти новые функции доверия будут равны:

$$g_1 = 0.316\eta_{\{v_2\}} + 0.009\eta_{\{v_3\}} + 0.011\eta_{\{v_5\}} + \\ + 0.091\eta_{\{v_2, v_3\}} + 0.116\eta_{\{v_2, v_5\}} + 0.081\eta_{\{v_3, v_5\}} + 0.376\eta_V,$$

$$g_2 = 0.171\eta_{\{v_1\}} + 0.223\eta_{\{v_3\}} + 0.019\eta_{\{v_4\}} + 0.275\eta_{\{v_5\}} + \\ + 0.074\eta_{\{v_3, v_4\}} + 0.08\eta_{\{v_3, v_5\}} + 0.063\eta_{\{v_4, v_5\}} + 0.095\eta_V,$$

$$g_3 = 0.218\eta_{\{v_2\}} + 0.204\eta_{\{v_4\}} + 0.285\eta_{\{v_5\}} + \\ + 0.072\eta_{\{v_1, v_2\}} + 0.071\eta_{\{v_2, v_5\}} + 0.062\eta_{\{v_4, v_5\}} + 0.088\eta_V,$$

$$g_4 = 0.014\eta_{\{v_2\}} + 0.277\eta_{\{v_3\}} + 0.248\eta_{\{v_5\}} + \\ + 0.072\eta_{\{v_2, v_3\}} + 0.09\eta_{\{v_2, v_5\}} + 0.098\eta_{\{v_3, v_5\}} + 0.201\eta_V,$$

$$g_5 = 0.261\eta_{\{v_2\}} + 0.28\eta_{\{v_3\}} + 0.175\eta_{\{v_4\}} + \\ + 0.072\eta_{\{v_1, v_2\}} + 0.055\eta_{\{v_2, v_3\}} + 0.072\eta_{\{v_3, v_4\}} + 0.085\eta_V.$$

Матрица пигнистических вероятностей агрегированных свидетельств  $Bet_k(\{v_i\})$  будет равна

$$Bet = \begin{pmatrix} 0.075 & 0.495 & 0.17 & 0.075 & 0.185 \\ 0.19 & 0.019 & 0.319 & 0.107 & 0.366 \\ 0.054 & 0.307 & 0.018 & 0.253 & 0.369 \\ 0.04 & 0.135 & 0.402 & 0.04 & 0.382 \\ 0.053 & 0.342 & 0.361 & 0.228 & 0.017 \end{pmatrix}.$$

Теперь вектор индексов взаимодействия  $q$  агрегированных свидетельств будет равен  $(0.412, 1.297, 1.269, 0.703, 1.318)$ . Таким образом, центральности вершин двухслойного графа теперь будут упорядочены следующим образом:  $x_1 \leq x_4 \leq x_3 \leq x_2 \leq x_5$ .

Можно доказать следующее утверждение о взаимодействиях вершин двухслойной сети без циклов, вычисленных указанным способом, в случае, когда центральности вычисляются с помощью 1-окрестности.

**Теорема 2.2.** Если  $G = (V, E, W)$  – граф без циклов с двумя слоями  $w^{(k)}(u, v)$ ,  $u, v \in V$ ,  $k = 1, 2$  и взаимодействие вершин вычисляется в 1-окрестности с помощью правила Демпстера, то

$$q(v) = \frac{1}{|W^{(1)}||W^{(2)}|} \sum_{u \in N^{(1)}(v)} \frac{w^{(1)}(u,v)w^{(2)}(u,v)}{1-K_u} + \frac{1}{|W^{(1)}|} \sum_{u \in N^{(1)}(v)} \frac{w^{(1)}(u,v)}{1-K_u} \left(1 - \frac{n^{(2)}(u)}{|W^{(2)}|}\right) +$$

$$+ \frac{1}{|W^{(2)}|} \sum_{u \in N^{(1)}(v)} \frac{w^{(2)}(u,v)}{1-K_u} \left(1 - \frac{n^{(1)}(u)}{|W^{(1)}|}\right) + B(G),$$

где  $n^{(k)}(u) = \sum_{x: e(u,x) \in E} w^{(k)}(u,x)$  – сумма весов всех ребер  $k$ -го слоя, инцидентных вершине  $u$ ,  $k=1,2$ ;  $|W^{(k)}| = \sum_{e(u,v) \in E} w^{(k)}(u,v)$  – сумма весов всех ребер  $k$ -го

слоя,  $k=1,2$ ;  $K_u = \frac{1}{|W^{(1)}||W^{(2)}|} \sum_{\substack{x \neq y, \\ e(u,x) \in V, \\ e(u,y) \in V}} w^{(1)}(u,x)w^{(1)}(u,y)$ ;

$$B(G) = \frac{1}{|V|} \sum_{u \in V} \frac{1}{1-K_u} \left(1 - \frac{n^{(1)}(u)}{|W^{(1)}|}\right) \left(1 - \frac{n^{(2)}(u)}{|W^{(2)}|}\right) - \text{константа, не зависящая от}$$

вершины  $v$ .

В случае применения правила Ягера формула для взаимодействия вершин упростится.

Теорема 2.3. Если  $G=(V,E,W)$  – граф без циклов с двумя слоями  $w^{(k)}(u,v)$ ,  $u,v \in V$ ,  $k=1,2$  и взаимодействие вершин вычисляется в 1-окрестности с помощью правила Ягера, то

$$q(v) = \frac{n^{(1)}(v)}{|W^{(1)}|} + \frac{n^{(2)}(v)}{|W^{(2)}|} +$$

$$+ \frac{1}{|W^{(1)}||W^{(2)}|} \sum_{u \in N^{(1)}(v)} \left( w^{(1)}(u,v)w^{(2)}(u,v) - n^{(1)}(u)w^{(2)}(u,v) - w^{(1)}(u,v)n^{(2)}(u) \right) + C(G),$$

$$\text{где } C(G) = \frac{1}{|V|} \sum_{u \in V} \left( K_u + \left(1 - \frac{n^{(1)}(u)}{|W^{(1)}|}\right) \left(1 - \frac{n^{(2)}(u)}{|W^{(2)}|}\right) \right) - \text{константа, не зависящая}$$

от вершины  $v$ .

Видно, что в этом случае взаимодействие представляет собой сумму средневзвешенных степенных центральностей в каждом слое и слагаемое, характеризующее взаимодействие ближайших к вершине  $v$  соседей с другими, соседними с ними вершинами.

Результаты этого подраздела отражены в [136].

## **2.7 Некоторые выводы**

В рамках работы построено обобщение разработанных ранее индексов дальних взаимодействий (LRIC), которые могут применяться в сетях, где связи между элементами приводят к их взаимозависимости. Определены наиболее важные элементы в сетях общей торговли, а также в сети взаимодействий террористических групп и показано отличие классических индексов центральности от предложенных индексов центральности LRIC, позволяющих учитывать особенности рассматриваемых сетей.

Предложенная в подразделе 2.2 модель имеет программную реализацию на языке Python и может быть скачана по адресу: <https://github.com/SergSHV/slrlic>.

Новые индексы центральности, CI, BI, PI, как и их обобщенный вариант были применены к различным реальным сетевым структурам. Отличительной особенностью индексов является учет, как индивидуального влияния, так и группового влияния, а также влияния внутри каждой группы стран. Влияние каждого из этих аспектов подробно описано для различных случаев изменения параметров индексов. Результаты показали, что изменения параметров индексов в некоторых случаях оказывают значительное влияние на рейтинги стран.

Данные индексы центральности были применены к сети международной миграции, а также к сети межстрановой торговли. Так, в сети международной торговли с помощью новых индексов были проанализированы все страны за четыре периода времени общей продолжительностью 15 лет. Результаты показывают, что наиболее влиятельными странами торговой сети являются Франция, Испания, Соединенные Штаты и Великобритания, а для миграционной сети – Соединенные Штаты, Германия, Великобритания и Южная Корея.

В данном отчёте также проведена предварительная работа с данными по экспорту и импорту нефти и газа, включающая в себя описание доступных данных в динамике и расчет базовых индексов центральности. В продолжение данной работы будут заполнены пропуски в данных, и добавлены значения за

2019 год. Также планируется расчет индексов SRIC (Short-Range Interactions) и LRIC (Long-Range Interaction Centralities) для сетей экспорта/импорта нефти и газа в двух состояниях в динамике для 2001 – 2019 годов, также подбор данных по экономическому развитию стран (темп роста ВВП), построение линейной регрессионной модели для данного показателя. Планируется анализ и подбор факторов, влияющих на объемы экспорта и импорта сырья в динамике, на основе этих данных проведение корреляционного анализа, построение моделей взаимосвязи, таких как линейная множественная регрессия и некоторые другие более продвинутые модели.

В данном исследовании по сетевому анализу была также представлена модель по оценке схожести слоёв многослойной сетевой структуры, которая в свою очередь может меняться во времени. Разработанная метрика основана на сравнении влияния вершины на её соседей на каждом уровне взаимодействия. Основным отличием данной работы от других метрик схожести графовых структур является то, что изначальная сеть нормализуется с учётом важности вершин, а также возможностью группового влияния, после чего сравнение сетей проводится только для ключевых вершин согласно некоторому показателю.

Результаты данной работы были применены к сети международного экспорта/импорта продовольствия, которая включает в себя 33 различных категории товаров за период с 1996 по 2017. Полученные результаты показали, что наиболее схожими с точки зрения международной торговли являются однородные группы товаров, а также те группы продуктов, которые одновременно экспортируют влиятельные страны в больших объёмах.

В этом разделе также представлены результаты по модификации гравитационной модели для анализа многосторонних торговых потоков. Модификация заключается во включении в модель отношения индексов центральности. Помимо представленных индексов в модель могут быть добавлены и дру-

гие меры центральностей. Результат этой работы можно использовать для более точных эконометрических оценок гравитационного уравнения торговли при рассмотрении международной торговой сети.

Наконец, в последнем подразделе представлены предварительные результаты по оцениванию центральностей многослойной сети методами теории функций доверия. Отличительной особенностью этого подхода является то, что агрегируются не центральности каждого слоя, а функции доверия локальных описаний сети в каждом слое. После чего вычисляются центральности по той информации, которую предоставляет новая (агрегированная) функция доверия.

### **3 Исследования по анализу текстов и кластерному анализу**

В этом разделе отчета описаны результаты исследований, полученные в ходе реализации проекта, по анализу текстов и кластерному анализу. Это исследование осуществлялось в следующих направлениях:

- разработка гибридного метода экономного подъема нечеткого тематического кластера в таксономии;
- разработка и исследование метода альтернирующих дихотомий для кластеризации данных;
- исследование государственной состоятельности в странах мира методами кластерного анализа.

#### **3.1 Разработка гибридного метода экономного подъема нечеткого тематического кластера в таксономии**

##### **3.1.1 Постановка задачи**

Автоматизация анализа и понимания текстов – одна из серьезных задач искусственного интеллекта. Несмотря на серьезные успехи, связанные с решением отдельных задач анализа текстов – категоризация, аннотирование, суммаризация, классификация и пр., общий анализ содержания, интерпретация текстовых коллекций остается серьезной проблемой для исследователей. Для облегчения интерпретации коллекций разнородных документов, получаемых по тем или иным правилам, фундаментальный характер приобретает необходимость сокращенного описания результатов поиска. Современные методы сокращения текстов основаны на “механистическом” вычленении тех или иных фрагментов текстов для помещения их в итоговое сокращенное представление. По нашему мнению, решение проблемы невозможно без попыток автоматизации когнитивной способности к обобщению, то есть замене совокупности “частных” понятий небольшим количеством обобщающих концепций. Это соответствует хорошо знакомому смыслу термина “обобщать” (см., например, Современный толковый словарь русского языка Ефремовой: “обобщать” – делать общим, соединять вместе; объединять).



Существующие подходы к вычислительному анализу структуры текстовых коллекций обычно не предполагают обобщения в качестве конкретно поставленной цели. Наиболее популярными инструментами для структурирования текстовых коллекций являются кластерный анализ и тематическое моделирование (topic modeling), описанное, например, в [137]. Оба способа включают признаки того же уровня детализации, что и отдельные слова или короткие фразы в текстах, таким образом, не давая решения задачи обобщения в качестве четко сформулированной цели. Поэтому, говоря формально, данные методы весьма спорно можно отнести к рассматриваемой задаче, однако они часто применяются для решения похожих проблем.

Тем не менее, иерархическая природа значений понятий и терминов отражается в потоке публикаций по анализу текста. Мы можем различать по крайней мере три направления, в контексте которых рассматривается вопрос обобщения. Прежде всего, следует упомянуть деятельность, связанную с развитием таксономий, в особенности, гипонимические / гипернимические отношения [138],[139]. Также следует упомянуть недавнюю статью [140], посвященную дополнению таксономии новыми темами исследований.

Другое направление является разделом классической деятельности по обобщению текста – автоматическое “аннотирование”, “реферирование” или “резюмирование”, позволяющее, в частности, получать краткие “выжимки” из более объемных текстов. Обычно, на практике “резюме” текста создаются с использованием механистических подходов к извлечению предложений. Тем не менее, отметим также подход для построения резюме как абстракций текстов путем объединения некоторых шаблонов, таких как триплеты субъект-глагол-объект (SVO) [141].

Еще одна сфера деятельности – это то, что можно назвать “операционным обобщением”. В этом направлении авторы используют обобщенные описания с использованием таксономических отношений между обобщенными состояниями и их составными частями для достижения

определенных целей, таких как улучшение характеристик поиска текста (см., например, [142], [143].

Мы будем использовать существующие таксономии для реализации идеи обобщения, не склоняясь ни к какому из описанных выше направлений. В соответствии со значением из Словаря современного русского языка, приведенным выше, будем понимать термин “обобщение” в смысле вывода общей концепции из отдельных частных. Мы предполагаем, что нам дается самая простая структура для такого вывода – таксономия предметной области. В этом понимании таксономия представляет собой корневое дерево над ключевыми понятиями предметной области, помечающими узлы дерева. Дерево отражает основные гипонимические / гипернимические отношения в области так, что узел А является родителем узла В, если отношение “В входит в А” истинно. Ситуация, которая требует особого внимания – это случай, когда мы должны обобщить нечеткий набор листов таксономии, представляя таким образом суть какого-то эмпирически наблюдаемого явления.

В частности, рассмотрим моделирование направлений новых исследований в области Науки о данных (Data Science). Data Science – это новая область компьютерных наук. Самая популярная таксономия в области компьютерных наук разработана вручную Всемирной ассоциацией вычислительной техники (Association for Computing Machinery – ACM), наиболее представительным органом в этой области; последний выпуск таксономии был опубликован в 2012 году под наименованием “Система классификации компьютерных наук ACM” (ACM Computing Classification System – ACM-CCS) [144]. Мы рассматриваем ее часть, связанную с Наукой о данных, расширенную с помощью добавления нескольких листьев (описанную в [145]). Добавление листьев сделано для того, чтобы лучше отразить тематику анализируемых научных статей рассматриваемой в дальнейшем коллекции [146].

В качестве коллекций документов мы рассмотрим: коллекцию научных статей, опубликованных за последний период в ряде журналов,

представляющих область Data Science (Наука данных), а также сформированную по поисковым запросам тематики Науки данных коллекцию из более широкого множества журналов.

На примере этих коллекций и таксономии ACM-CCS мы собираемся найти кластеры тем максимально “тесным” образом, каждый из которых представляет тенденцию развития, отраженную в коллекции. В этом случае естественно предположить, что кластеры будут нечеткими, в соответствии с нечеткой природой семантики научных статей. Основной проблемой является интерпретация такого нечеткого кластера. Как уже указывалось, обычно интерпретация кластеров осуществляется с использованием понятий того же уровня гранулярности (признаки, ключевые слова), которые используются при построении кластеров. Мы предлагаем использовать понятия укрупненной гранулярности, для чего воспользоваться таксономией предметной области. Конкретно, речь идет об обобщении понятий таксономии, соответствующих полученному кластеру.

### 3.1.2 Экономный подъем нечеткого тематического кластера в таксономии: модель и метод

Говоря математическим языком, таксономия – это корневое дерево, чьи узлы аннотированы темами таксономии. Мы будем рассматривать следующую задачу: дано нечеткое множество  $S$  из листьев таксономии, необходимо найти узел  $t(S)$  более высокого ранга в таксономии, который покрывает множество  $S$  наиболее “тесным” образом. Такая задача “подъема” – математическая постановка человеческой способности делать обобщение, то есть, процесса формирования “концептуальной” формы представленного явления, в данном случае, нечеткого множества листьев дерева.

Задача не так проста, как могло бы показаться. Для краткости, рассмотрим четкое множество  $S$ , представленное пятью черными прямоугольниками на фрагменте дерева на рисунке 3.1. Рисунок 3.2 иллюстрирует ситуацию, при которой множество поднято в корень, что

показано закрашиванием соответствующего ему прямоугольника, как и его потомков.

Если мы допустим, что множество  $S$  может быть обобщено корнем, это приведет к тому, что корнем будет покрыто и четыре белых прямоугольника, и, таким образом, попадет в то же обобщение, что и  $S$ , при том, что они не принадлежат множеству  $S$ . Такую ситуацию будем далее называть появлением пробелов. Очевидно, что вариант подъема, при котором возникают пробелы, должен штрафоваться моделью подъема. В целом, число концептуальных элементов, введенных для обобщения  $S$ , здесь равно 1 головной вершине (то есть корню), который мы теперь присвоили переменной  $S$ , и 4 пробела, образовавшиеся из-за топологии дерева. При этом за подобный подъем, как уже было сказано, должно быть назначено определенное значение штрафа. Другой возможный вариант подъема показан на рисунке 3.3: здесь заданное множество поднимается только до корня левой ветви дерева. Мы можем видеть, что число пробелов резко сократилось – до 1. Однако возник еще один тип неточности подъема: черный прямоугольник справа, принадлежащий  $S$ , не покрыт корнем левой ветви, в которую отображается множество  $S$ . Этот тип ошибки будем называть выбросом. При этом варианте подъема появляются три новых объекта: одна головная вершина, один выброс и один пробел. Это меньше, чем количество объектов, появившихся при поднятии множества к корню (одна головная вершина и четыре пробела, то есть совокупно – пять), что делает его более предпочтительным. Конечно, этот вывод верен только в том случае, если относительный штраф за один выброс меньше суммарного штрафа за три пробела.



Рисунок 3.1 – Фрагмент таксономии, черные прямоугольники соответствуют элементам, принадлежащим множеству

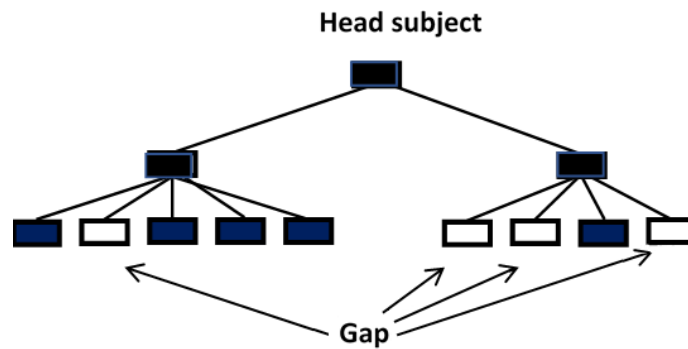


Рисунок 3.2 – Обобщение нечеткого множества запроса, показанного на рисунке 3.1 с помощью его отображения в корень дерева. Цена этого обобщения - четыре пробела (gap)

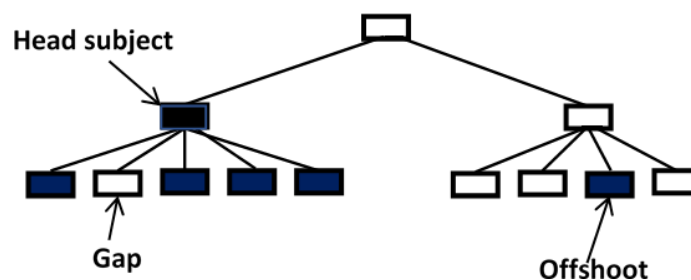


Рисунок 3.3 – Обобщение нечеткого множества запроса, показанного на рисунке 3.1 с помощью его отображения в корень левого поддерева. Цена такого обобщения - один пробел (gap) и один выброс (offshoot)

Нас интересует, можно ли обобщить нечеткое множество  $S$  в узел  $t$  принадлежащий более высокому рангу таксономии, так, чтобы можно было считать, что  $S$  попадает в покрытие, охватываемое узлом  $t$ . Если это будет сделано напрямую, например, путем выбора последнего общего предка для узлов в  $S$  получение желаемого охвата будет маловероятным. Чтобы сделать процесс обобщения более адекватным, следует допустить наличие в итоговом обобщении не только пробелов, но и выбросов, являющихся листьями, принадлежащими обобщению, но не исходному множеству.

Цель нахождения интерпретируемого обобщения для  $S$  в рамках таксономии может быть формализована как нахождение одного или нескольких “головных тем” для покрытия  $S$  с минимальным количеством новых элементов всех типов, получившихся в процессе обобщения: главных

вершин, пробелов и выбросов. Такая постановка цели реализует принцип максимальной экономии (Maximum Parsimony, MP) при описании данного явления, популярный в некоторых областях, таких как биоинформатика.

Рассмотрим корневое дерево  $T$ , представляющее иерархическую таксономию, такую, что его узлы снабжены ключевыми фразами, обозначающими различные понятия. Обозначим множество его листьев через  $I$ . Соотношение между узлами в иерархии обычно выражается с использованием генеалогических терминов: каждый узел  $t \in T$  называется родителем узлов, непосредственно ответвляющихся от  $t$ , его потомков. Будем использовать  $\chi(t)$  для обозначения множества потомков этого узла. Предполагается, что каждый внутренний узел  $t \in T - I$ , то есть узел, не являющийся листом таксономии, соответствует понятию, обобщающему темы, соответствующие листьям  $I(t)$ , выходящим из  $t$ , то есть, листья поддерева  $T(t)$  имеют общий корень в узле  $t$ , называемый кластером листьев.

Нечеткое множество на  $I$  – это отображение  $u$  из множества  $I$  в множество неотрицательных действительных чисел, которое присваивает значение функции принадлежности (или поддержку)  $u(i) \geq 0$  каждому узлу  $i \in I$ . Мы называем множество, где  $S_u \subset I$ , где  $S_u = \{i \in I: u(i) > 0\}$ , поддержкой  $u$ . Никаких других предположений о виде функции  $u$  не делается. Отметим только то, что для удобства целесообразно положить такое ограничение, чтобы ее значение не превышало единицу. Классические, или четкие множества соответствуют бинарным функциям принадлежности – таким, что  $u(i)=1$ , если  $i \in S_u$  и  $u(i)=0$  иначе.

Рассматривая нечеткое множество запроса  $u$ , определенное на листьях  $I$  дерева  $T$ , можно считать  $u$  (возможно, сделанной с погрешностями) проекцией понятия более высокого ранга, “головной темы”, на соответствующий кластер листьев. В этом предположении должен существовать головной узел  $h$  среди внутренних узлов дерева  $T$  – такой, что его кластер листьев  $I(h)$  в

определенной степени совпадает (с небольшими ошибками) с  $S_u$ . Этот головной узел – обобщение  $u$ , которое и необходимо найти.

Возможны два типа ошибок, связанных с поиском головного узла, если он точно не покрывает поддержку, которые представляют собой ложные срабатывания и ложные отрицания, уже упомянутые ранее как пробелы (gaps) и выбросы (offshoots) – соответственно, иллюстрируются на рисунках 3.2 и 3.3. Общее количество головных узлов, пробелов и выбросов должно быть как можно меньше. Для этого мы вводим штраф за каждый из этих элементов. Предполагая для простоты, что все черные прямоугольники на рисунке 3.1 имеют значения функции принадлежности, равные единице, можно легко увидеть, например, что общее значение штрафа, например, для подъема на рисунке 3 равно  $1 + \gamma + \lambda$ , где  $\gamma$ ,  $\lambda$  и 1 – штрафы за выброс, пробел и головную тему соответственно.

Рассмотрим узел-кандидат  $h$  в  $T$  и его значение принадлежности в нечетком множестве  $u$ .  $h$ -пробел – это узел  $g$  из  $T(h)$ , отличный от  $h$ , в котором произошла потеря значения, то есть  $g$  является максимальным  $u$ -иррелевантным узлом в том смысле, что его родитель не является  $u$ -иррелевантным. И наоборот, установление узла  $h$  в качестве головного можно рассматривать как приобретение значения  $u$  в узле. Множество всех  $h$ -выбросов обозначим как  $G(h)$ .

Чтобы правильно определить идею использования пробелов, введем понятие  $u$ -иррелевантных узлов. Узел называется  $u$ -иррелевантным, если его кластер листьев  $I(t)$  не пересекается с поддержкой  $\overline{[u]}$ . Очевидно, что если узел принадлежит  $u$ , все его потомки также принадлежат  $u$ .

$h$ -выброс – это лист  $i \in S_u$ , который не покрывается  $h$ , то есть  $i \notin I(h)$ . Множество всех  $h$ -выбросов – это  $S_u - I(h)$ .

Если задано нечеткое тематическое множество  $u$  над  $I$ , множество узлов  $H$  будет называться  $u$ -покрытием, если: (а)  $H$  покрывает  $S_u$ , то есть  $S_u \subseteq \bigcup_{h \in H} I(h)$ , и (б) узлы в  $H$  не связаны, то есть  $I(h) \cap I(h') = \emptyset$  для любых  $h, h' \in H$ , таких, что  $h \neq h'$ . Внутренние узлы  $H$  будут называться головными

темами, а листовые узлы – выбросами, поэтому множество выбросов в  $H$  равно  $H \cap I$ . Множество пробелов в  $H$  является объединением  $G(h)$  по всем головным темам  $h \in H - I$ .

Определим штрафную функцию  $p(H)$  для  $u$ -покрытия  $H$  следующим образом (формула (3.1)):

$$p(H) = \sum_{h \in H - I} u(h) + \sum_{h \in H - I} \sum_{g \in G(h)} \lambda v(g) + \sum_{h \in H \cap I} \gamma u(h). \quad (3.1)$$

Задача, которую мы решаем, состоит в том, чтобы найти такое  $u$ -покрытие  $H$ , которое минимизирует штрафную функцию  $p(H)$ . Такое  $u$ -покрытие и будет наиболее оптимальным обобщением нечеткого множества  $u$ .

Перед применением алгоритма для минимизации суммарного штрафа необходимо выполнить предварительное преобразование дерева. Для этого необходимо сделать следующие шаги. Из дерева нужно удалить все немаксимальные  $u$ -иррелевантные узлы, то есть потомки пробелов, а также найти множества  $G(t)$  и значения суммарной важности пробелов  $V(t) = \sum_{g \in G(t)} v(g)$  для всех внутренних узлов  $t$ . Отметим, что все элементы  $S_u$  находятся в множестве листьев обрезанного дерева, а остальные листья обрезанного дерева – это именно пробелы.

Теперь мы можем применить алгоритм подъема ПарГеНМ. Для каждого узла  $t$  алгоритм ПарГеНМ вычисляет два набора –  $H(t)$  и  $L(t)$ , содержащие такие узлы в  $T(t)$ , в которых, соответственно, происходят приобретения и потери головных узлов (включая выбросы). Соответствующий штраф обозначается  $p(t)$  и вычисляется так, как описано ниже.

Предположение алгоритма состоит в том, что после потери узла не может быть последующего его приобретения. Следовательно,  $H(t)$  и  $L(t)$  вычисляются при условии, что головной узел не был ни приобретен (ни потерян) ни у одного из предков  $t$ . Алгоритм ПарГеНМ рекурсивно вычисляет  $H(t)$ ,  $L(t)$  и  $p(t)$  из соответствующих значений для дочерних узлов в  $\chi(t)$ . Для



каждого листового узла, который не принадлежит  $S_u$ , мы определяем как  $L(\cdot)$ , так и  $H(\cdot)$  пустыми, а штраф назначается равным нулю. Для каждого листового узла, который находится в  $S_u$ , набор  $L(\cdot)$  устанавливается пустым, тогда как  $H(\cdot)$ , содержит только листовой узел, и штраф определяется как произведение зафиксированного параметра  $\gamma$  на значение его принадлежности. Чтобы вычислить  $L(t)$  и  $H(t)$  для любого внутреннего узла  $t$  далее устанавливаем, какое событие имело место быть: (а) головной узел был приобретен в  $t$ ; (б) головной узел не был приобретен в  $t$ .

В случае (а) множества  $H(\cdot)$  и  $L(\cdot)$  у его потомков далее оказываются не нужными. В этом случае,  $H(t)$ ,  $L(t)$  и  $p(t)$  определяются как (формула (3.2.)):

$$\begin{cases} H(t) = \{t\}, \\ L(t) = G(t), \\ p(t) = u(t) + \lambda V(t). \end{cases} \quad (3.2)$$

В случае (б) множества  $H(t)$  и  $L(t)$  являются просто объединениями соответствующих множеств из их потомков, а  $p(t)$  равна сумме всех штрафов из потомков:

$$\begin{cases} H(t) = \bigcup_{w \in \chi(t)} H(w), \\ L(t) = \bigcup_{w \in \chi(t)} L(w), \\ p(t) = \sum_{w \in \chi(t)} p(w). \end{cases} \quad (3.3)$$

Чтобы получить оптимальный подъем, из этих случаев выбирается тот, где достигается меньшее значение  $p(t)$ . В ситуации, когда оба случая дают одинаковые значения  $p(t)$ , мы можем выбрать любой из них произвольно – в формулировке алгоритма ниже, мы выбрали (а). Выходные данные алгоритма включают значения рассчитываемых множеств в корне, а именно:  $H$  – множество головных узлов и выбросов,  $L$  – множество пробелов и  $p$  – назначенное значение штрафа. Мы доказали, что алгоритм ПарГеНМ действительно приводит к глобально оптимальному обобщению [146].

### 3.1.3 Изучение направлений исследований на основе коллекций научных публикаций с помощью таксономии предметной области

Продemonстрируем применение алгоритма ПарГеНМ для задачи анализа коллекций на примере работы с двумя коллекциями научных публикаций. Можно предложить два способа анализа: первый – эмпирический, второй – теоретический. Первый способ предлагает определить категории, присутствующие в коллекции, из самих текстов, второй – на основе знаний о предметной области. Первый подход, например, может быть хорошо проиллюстрирован тематическим моделированием на основе LDA (в частности, [137]); второй подход – применением таксономии, разрабатываемой экспертами, такой как ACM-CCS (подход описан, например, в [147]).

Для анализа коллекции мы последовательно применяем шаги, перечисленные ниже:

- подготовка коллекции научных статей;
- подготовка таксономии рассматриваемой предметной области;
- расчет матрицы значений релевантности между тематическими листьями таксономии и публикациями из коллекции статей;
- поиск нечетких кластеров по матрице значений релевантности;
- подъем полученных кластеров в таксономии для их концептуализации с помощью обобщения;
- построение выводов из обобщений.

Каждый из этих пунктов далее будет рассмотрен в отдельном подразделе.

#### 3.1.3.1 Подготовка коллекции научных публикаций №1

Благодаря доступу к библиотеке журналов издательства Springer нам удалось загрузить коллекцию из 17685 научных статей вместе с их аннотациями, опубликованных в 17 журналах, связанных с Наукой о данных,

за 20 лет: с 1998 по 2017 год. В качестве текстовой коллекции №1 были рассмотрены аннотации этих статей.

### 3.1.3.2 Таксономия DST

Таксономия – это форма инженерии знаний, которая в последние годы становится все более популярной. Наиболее известны таксономии, созданные в рамках проекта Genome Ontology (GO) [148] в области биоинформатики, проекта SNOMED CT для здравоохранения и медицины [149] и тому подобное.

Математически таксономия – это корневое дерево, иерархия, все узлы которой помечены основными понятиями предметной области. Иерархия соответствует отношению включения: тот факт, что узел А является родителем В, означает, что В является частью или частным случаем А. Важной характеристикой корневого дерева является то, что у каждого узла есть только один родительский элемент.

Существует два основных подхода к разработке таксономии домена: автоматический и ручной. Последний, безусловно, в настоящее время является гораздо более зрелым и развитым. Тем не менее, даже этот подход страдает от недостатков, которые сводятся к тому, что решения по разработке таксономии, касающиеся используемых структур классификации, процедур и описательных основ, как правило, недостаточно хорошо описаны и мотивированы. Автоматический подход может использовать множество цифровых ресурсов и методов семантического анализа. На данный момент чаще предпочитают таксономии, разрабатываемые вручную представительным органом специалистов. Действительно, такая таксономия не зависит от чисто эмпирических данных, используемых автоматическими методами.

Более того, ручная таксономия обычно уравнивает теоретические наработки и практический опыт, накопленный в сообществе, представленном органом, который ее разрабатывает. Такова, например, Система классификации компьютерных наук Всемирной ассоциации вычислительной

техники ACM (ACM-CCS 2012). Выбранное нами ее подмножество Data Science – Наука о данных, включает такие области, как машинное обучение, интеллектуальный анализ данных, анализ данных и так далее. Мы берем ту часть таксономии ACM-CCS 2012, которая связана с Data Science, и добавляем несколько листов, связанных с более поздними разработками Data Science. Рассматриваемая таксономия целиком опубликована в [145], [146]. Темы высших рангов таксономии представлены в Таблице 3.1.

Таблица 3.1 – Подмножество таксономии Системы классификации компьютерных наук Всемирной ассоциации вычислительной техники ACM (ACM-CCS) 2012, касающееся Науки о данных; несколько тем самых высоких рангов

Subject index	Subject name
1.	Theory of computation
1.1.	Theory and algorithms for application domains
2.	Mathematics of computing
2.1.	Probability and statistics
3.	Information systems
3.1.	Data management systems
3.2.	Information systems applications
3.3.	World Wide Web
3.4.	Information retrieval
4.	Human-centered computing
4.1.	Visualization
5.	Computing methodologies
5.1.	Artificial intelligence
5.2.	Machine learning

### 3.1.3.3 Вычисление степени релевантности между текстами и темами таксономии

Для вычисления степеней релевантности тем таксономии и текстовых документов коллекции будем использовать метод АСД, описанный, например, в [150]. Как уже было сказано, существует большое количество его валидаций на различных коллекциях, показавших его эффективность. Данный метод не требует предобработки текстов и учитывает порядок слов, в отличие от большинства популярных подходов.

Чтобы уменьшить влияние часто встречающихся общих терминов, функция оценки модифицируется введением пятикратного уменьшения веса

специальных стоп-слов. Список стоп-слов включает: “learning, analysis, data, method” и несколько постфиксов: “s/es, ing, tion”.

#### 3.1.3.4 Определение и вычисление нечетких кластеров тем таксономии

Группы тем, которые мы планируем получить, должны отражать одновременное возникновение тем: чем больше число текстов, к которым относятся обе темы  $t$  и  $t'$ , тем больше взаимосвязь между  $t$  и  $t'$ , тем больше вероятность того, что темы  $t$  и  $t'$  попадут в один кластер.

Было протестировано несколько популярных алгоритмов кластеризации: K-Means, DBSCAN и классический спектральный метод, которые не показали удовлетворительных результатов, ввиду того, что не дали интерпретируемых кластеров. Поэтому мы представляем здесь результаты, полученные с помощью алгоритма FADDIS, описанного в [147], и разработанного специально для поиска нечетких тематических кластеров. Этот алгоритм реализует следующие предположения, относящиеся к рассматриваемой задаче:

- [LN] нормализация Лапласа: преобразование матрицы похожести, выполняемое для того, чтобы сделать кластерную структуру четче;
- [AA] аддитивность: тематические кластеры за текстами являются аддитивными, поэтому значения сходства представляют собой суммы вкладов по различным скрытым темам;
- [AN] неполнота: кластеры не обязательно охватывают все ключевые фразы, имеющиеся в наличии, поскольку рассматриваемая текстовая коллекция может не иметь отношения к некоторым из них.

##### 3.1.3.4.1 Оценивание степени близости тем

Основываясь на матрице релевантности ключевых фраз и документов коллекции  $R$ , можно получить матрицу близости между ключевыми фразами –  $A$ , которая, говоря иными словами, представляет собой матрицу “совместной релевантности” ключевых фраз в соответствии со структурой текстовой коллекции. Элементы этой матрицы, то есть значения сходства  $a_{tt'}$ , между темами  $t$  и  $t'$  можно вычислить как скалярное произведение векторов оценок

$r_t = (r_{tv})$  и  $r_{t'} = (r_{t'v})$ , где  $v=1,2,\dots,V=17685$ . Скалярное произведение можно модифицировать, используя естественные весовые коэффициенты, присваиваемые документам в коллекции. Вес текстового документа  $v$  определяется как отношение количества всех относящихся к нему тем  $n_v$  и  $n_{max}$ , максимуму  $n_v$  для всех  $v = 1,2,\dots,V$ . Тема считается относящейся к  $v$ , если показатель ее релевантности документу превышает 0.2 (этот порог найден экспериментально, показано в [150]).

#### 3.1.3.4.2 Аддитивная нечеткая спектральная кластеризация

Обозначим множество листовых тем через  $T$  и предположим, что нечеткий кластер над  $T$  представлен вектором принадлежности  $\vec{u} = u_t, t \in T$ , таким что  $0 \leq u_t \leq 1$  для всех  $t \in T$ , а также пусть интенсивность  $\mu > 0$  – масштабирующий коэффициент, связывающий значения принадлежности со значениями близости. Для  $T$ , представляющего собой набор тем исследований, и  $\vec{u} = u_t, t \in T$ , вектора значений принадлежности, представляющего собой семантическую подструктуру совокупности рассматриваемых публикаций, произведение  $(\mu u_t)(\mu u_{t'}) = \mu^2$  можно рассматривать как вклад направления исследований, представленного рассматриваемым кластером в общую оценку близости  $a_{tt'}$  между темами  $t$  и  $t'$ .

Модель аддитивной нечеткой кластеризации, предложенная в [151], утверждает, что значения элементов матрицы близости тем  $A$  можно рассматривать как результат аддитивного вклада нечетких кластеров  $K$ , вплоть до ошибок, которые должны быть минимизированы (формула 3(4):

$$a_{tt'} = \sum_{k=1}^K \mu_k^2 u_{kt} u_{kt'} + e_{tt'}, \quad (3.4)$$

где  $\vec{u}_k = (u_{kt})$  – вектор принадлежности кластера  $k$ , и  $\mu_k$  – его интенсивность.

Эти предположения требуют, чтобы кластеры были извлечены в соответствии с аддитивной моделью. Метод, разработанный в [151],

FADDIS, находит кластеры последовательно, один за другим, что соответствует сформулированным выше предположениям, минимизируя:

$$E = \sum_{t,t' \in T} (w_{tt'} - \chi u_t u_{t'})^2. \quad (3.5)$$

В статье [151] приведено большое количество теоретических и экспериментальных результатов, призванных продемонстрировать тот факт, что FADDIS конкурентоспособен по сравнению с другими подходами нечеткой кластеризации и имеет ряд преимуществ.

### 3.1.3.4.3 Тематические кластеры FADDIS

После вычисления матрицы близости тем размера  $317 \times 317$  (по количеству листовых тем таксономии), применения псевдообратного преобразования Лапласа, и кластеризации методом FADDIS, мы последовательно получили 6 кластеров, три из которых кажутся особенно однородными. Основываясь на их содержимом, 'Learning', 'Retrieval' и 'Clustering', обозначим их L, R и C соответственно. Эти кластеры представлены в Таблице 3.2.

Таблица 3.2 – Кластеры L, R, C: темы с наибольшими значениями функции принадлежности

Cluster L			Cluster R			Cluster C		
u(t)	Code	Topic	u(t)	Code	Topic	u(t)	Code	Topic
0.300	5.2.3.8.	rule learning	0.211	3.4.2.1.	query representation	0.327	3.2.1.4.7.	biclustering
0.282	5.2.2.1.	batch learning	0.207	5.1.3.2.1.	image representations	0.286	3.2.1.4.3.	fuzzy clustering
0.276	5.2.1.1.2.	learning to rank	0.194	5.1.3.2.2.	shape representations	0.248	3.2.1.4.2.	consensus clustering
0.217	1.1.1.11.	query learning	0.194	5.2.3.6.2.1.	tensor representation	0.220	3.2.1.4.6.	conceptual clustering
0.216	5.2.1.3.3.	apprenticeship learning	0.191	5.2.3.3.3.2.	fuzzy representation	0.192	5.2.4.3.1.	spectral clustering
0.213	1.1.1.10.	models of learning	0.187	3.1.1.5.3.	data provenance	0.187	3.2.1.4.1.	massive data clustering
0.203	5.2.1.3.5.	Adversarial learning	0.173	2.1.1.5.	equational models	0.159	3.2.1.7.3.	graph based conceptual clustering

### 3.1.3.5 Результаты подъема кластеров L, R, и C в таксономии DST

Полученные кластеры были подняты в таксономии DST с использованием алгоритма ПарГеНМ, со штрафом за пробел  $\lambda=0.1$  и выброс  $\gamma=0.9$ .

Результаты подъема кластера L показаны на рисунке 3.4. Кластер получил три головных темы: machine learning, machine learning theory, and learning to rank. Они представляют структуру общего понимания “Learning” в соответствии с нашей текстовой коллекцией. Список полученных пробелов является менее представительным, отражая, вероятно, относительно скромный охват этих областей публикациями в коллекции (пробелы показаны в Таблице 3.3).

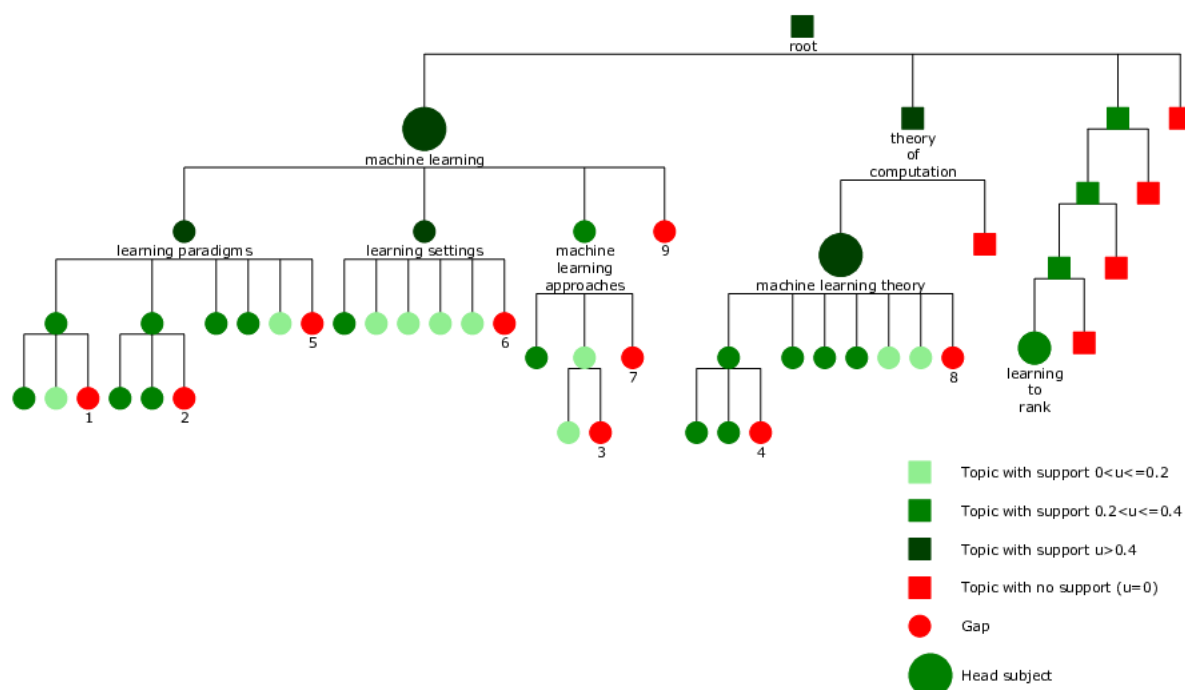


Рисунок 3.4 – Результаты подъема кластера L: Learning. Пробелы пронумерованы, (Таблица 3.3)



Таблица 3.3 – Пробелы, полученные в процессе подъема кластера L

Номер	Темы
1	ranking, supervised learning by classification, structured outputs
2	sequential decision making in practice, inverse reinforcement learning in practice
3	statistical relational learning
4	sequential decision making, inverse reinforcement learning
5	unsupervised learning
6	learning from demonstrations, kernel approach
7	classification and regression trees, kernel methods, neural networks, learning in probabilistic graphical models, learning linear models, factorization methods, markov decision processes, stochastic games, learning latent representations, multiresolution, support vector machines
8	sample complexity and generalization bounds, boolean function learning, kernel methods, boosting, bayesian analysis, inductive inference, structured prediction, markov decision processes, regret bounds
9	machine learning algorithms

Аналогичные комментарии можно сделать и в отношении результатов подъема кластера R: Retrieval (рисунок 3.5). Получены следующие головные темы: Information Systems and Computer Vision, они показывают структуру публикаций по тематике “Retrieval” в рассматриваемой коллекции.

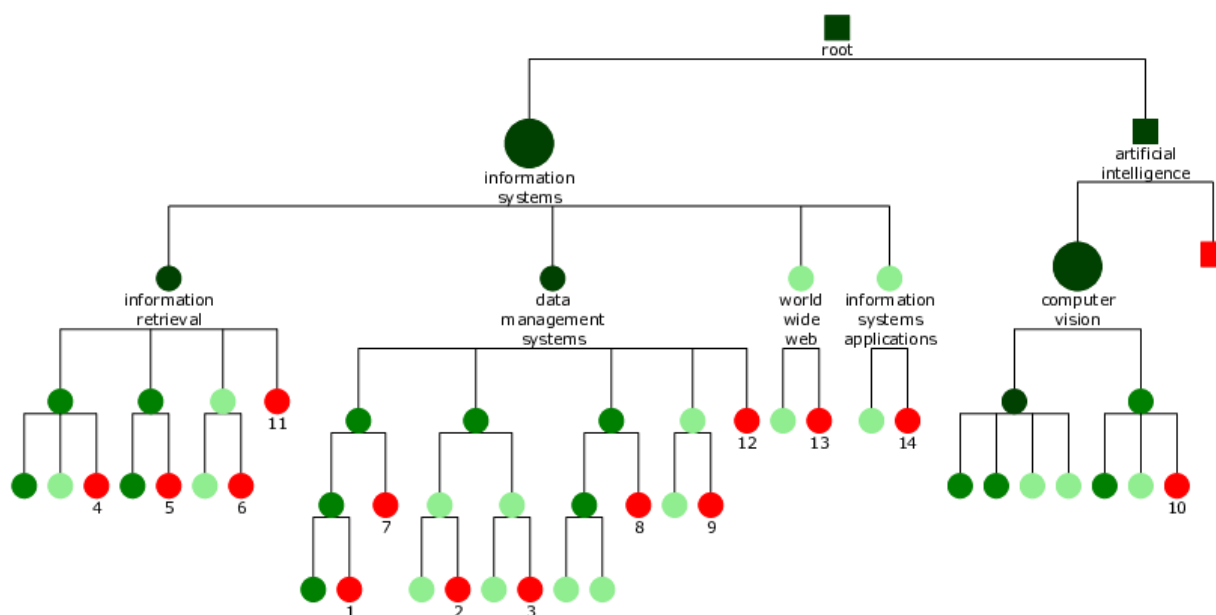


Рисунок 3.5 – Результаты подъема кластера R. Пробелы пронумерованы. Обозначения к рисунку приведены в легенде рисунка 3.4

Мы не приводим иллюстрации подъема кластера C, так как соответствующий фрагмент таксономии оказался слишком большим, а

результаты - слишком фрагментарны. Здесь было получено 16(!) головных тем: clustering, graph based conceptual clustering, trajectory clustering, clustering and classification, unsupervised learning and clustering, spectral methods, document filtering, language models, music retrieval, collaborative search, database views, stream management, database recovery, mapreduce languages, logic and databases, language resources. Как можно видеть, ключевые темы кластеризации дополняются методами и приложениями – это показывает, что постоянно возрастающая роль кластеризации в Науке о данных, возможно, должна быть лучше отражена в таксономии.

### 3.1.3.6 Выводы, полученные по итогам анализа коллекции №1

Можно видеть, что тематические кластеры, найденные в рассматриваемой коллекции научных публикаций, составляют области наиболее актуальных разработок. Три рассматриваемых кластера тесно связаны со следующими процессами:

- теоретические и методологические исследования в машинном обучении, а также слияние машинного обучения с задачами ранжирования;
- представление различных типов данных для поиска информации и их объединение с визуальными данными, а также их семантикой;
- различные типы кластеризации в разных отраслях таксономии, связанные с разнообразными приложениями.

В частности, из головных тем кластера “Learning” хорошо заметно (рисунок 3.4 и комментарии к нему), что основная деятельность здесь все еще сосредоточена на теории и методах, а не на приложениях. Хорошей новостью является то, что области машинного обучения, ранее сфокусированные в основном на задачах обучения на подмножествах и разбиениях, в настоящее время расширяются в направлении обучения ранжированию и рейтингованию. Конечно, остается много областей, которые еще предстоит исследовать более подробно: их можно увидеть в списке пробелов в Таблице 3.3.

Переходя к анализу результатов поднятия кластера R “Retrieval” (рисунок 3.5), мы можем четко видеть тенденции современного этапа развития

этого раздела Науки о данных. Вместо того, чтобы относить термин “информация” только к текстам, как это было на предыдущих этапах развития информационных наук, в наши дни анализ визуальных представлений становится частью концепции информации. Однако есть одна загвоздка. В отличие от многоуровневой гранулярности представлений текстов, разработанной человечеством в течение тысячелетий в процессе общения через естественные языки, не существует сопоставимой иерархии для изображений. Можно предположить, что элементы кластера R, относящиеся к сегментации изображений и видео, а также элементы, относящиеся к системам управления данными – это те элементы, которые будут положены в основу будущей многоуровневой системы анализа изображений и видео. Это – потенциальное направление для будущих разработок, оно совершенно четко видно из рисунка 3.5.

Что же касается кластера C “Clustering” с его 16 головными темами, можно сделать вывод, что, возможно, уже наступил тот момент времени (или же такой момент должен наступить уже очень скоро), когда тема кластеризации должна быть поднята на более высокий уровень таксономии, чтобы охватить все эти головные узлы. В начале века Data Science, несколько десятилетий назад, кластеризация, как правило, считалась в значительной степени вспомогательной частью машинного обучения, обучением без учителя. Возможно, вскоре мы увидим новую таксономию Data Science, в которой кластеризация – это не просто вспомогательный инструмент, а скорее, модель эмпирической классификации, значительная часть инженерии знаний. Обсуждая роль классификации как явления инженерии знаний, мы сталкиваемся с тремя общепринятыми аспектами классификации:

- структурирование явлений;
- связывание различных аспектов явлений друг с другом;
- формирование и хранение знаний о явлениях.

Каждый из них может представлять отдельное направление исследований в области инженерии знаний.

### 3.1.3.7 Подготовка коллекции №2

В качестве исходных данных были использована коллекция из 68 933 аннотаций научных статей вместе с ключевыми словами, названием журнала и датой публикации. Исходные текстовые данные были получены следующим способом:

- рассматривались две базы данных журналов: Springer и Elsevier, доступные через электронную библиотеку ВШЭ;
- к интерфейсу электронной библиотеки с уточнением категории статей «компьютерные науки» были сделаны следующие запросы: clustering, machine learning, neural networks, algorithm, classification, information retrieval, natural language processing, software, computing, pattern recognition, deep learning, probabilistic, artificial intelligence, support vector, bayesian, regression, search engine;
- все найденные через систему статьи с непустой аннотацией и множеством ключевых словосочетаний загружены специально разработанным краулером.

Далее была проведена фильтрация коллекции (в скобках приведено количество статей, оставшихся на текущем шаге):

- выбраны 80 релевантных тематике Науки данных журналов (26 826);
- удалены статьи, длина аннотации к которым меньше 100 символов (26 799).

После всех преобразований в коллекции осталось 26 799 статей из 80 журналов за период с 1971 по 2019 год.

### 3.1.3.8 Таксономия

Для анализа была использована обновленная таксономия Науки данных (DST). В этой модифицированной таксономии наук о данных присутствует 456 вершин, из которых листьями являются 353 вершины (на 36 листьев больше, чем в таксономии из предыдущего расчета). Максимальная глубина таксономии составляла все те же 6 уровней.

С помощью метода АСД была получена матрица значений релевантности текстов коллекции и листовых тем таксономии размером 26799 на 353, а затем – матрица корелевантности.

3.1.3.9 Тематические кластеры FADDIS, полученные для коллекции, и их обобщение

С помощью метода FADDIS получено 35 нечетких кластеров на множестве тем. В качестве критерия останова использовано условие положительности веса кластера. Для исключения шумовых элементов из кластеров удалены те темы, значение принадлежности которых оказалось меньше 0.1.

Полученные кластеры обобщены с помощью алгоритма ПарГеНМ. После обобщения всех 35 кластеров были получены следующие результаты:

- 15 кластеров не получили ни одного головного понятия. Элементы этих кластеров очень слабо связаны между собой. Одним из объяснений их появления является наличие некоторого количества шума в исходной коллекции текстов, а так же использование метода АСД, который учитывает исключительно синтаксическую схожесть текстов (к примеру, оценка схожести АСД слов «morphology» и «onthology» высока, несмотря на то что эти термины принадлежат совершенно разным областям науки данных).

- 7 кластеров имеют меньше десяти листовых элементов. Эти кластеры, в основном, имеют одно или два головных понятия и несколько выбросов со значением функций принадлежности  $u < 0.4$ . Такие кластеры не представляют большого интереса, т.к. объединяют слишком малое множество вершин таксономии.

- 7 кластеров являются легко интерпретируемыми.

Примеры наиболее репрезентативных и интерпретируемых обобщений кластеров приведены на рисунках 3.6 и 3.7. В Таблице 3.4 представлена подробная информация по полученным кластерам.

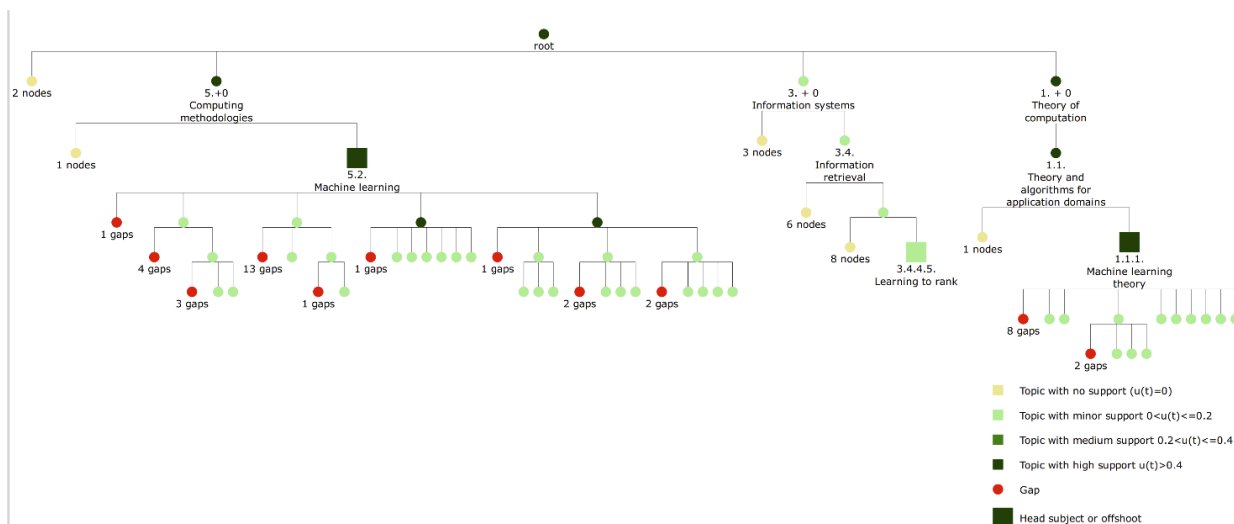


Рисунок 3.6 – Схематичное изображение результатов подъема кластера Learning на дереве таксономии

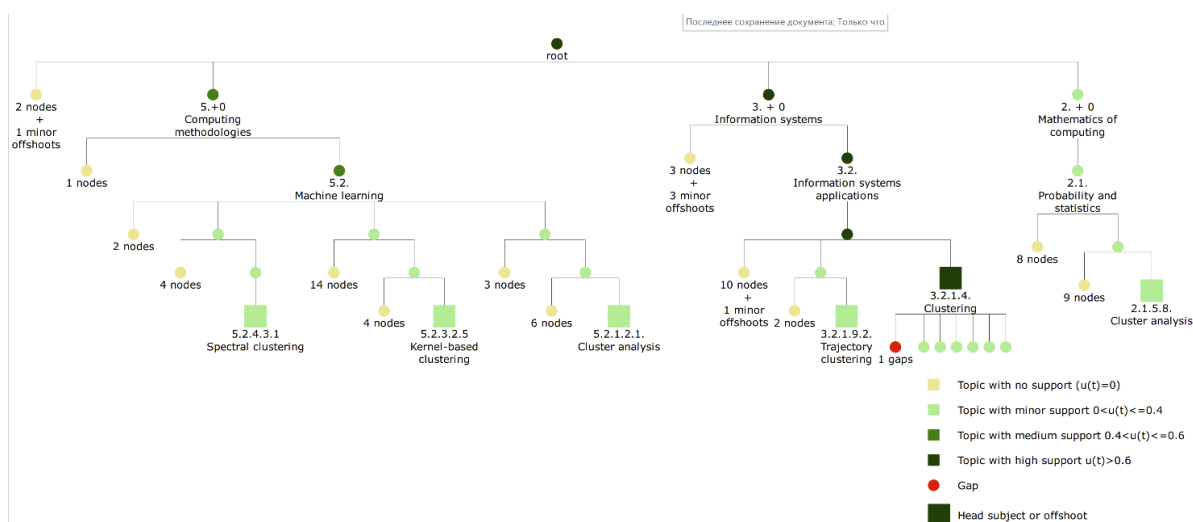


Рисунок 3.7 – Схематичное изображение результатов подъема кластера Clustering на дереве таксономии

Таблица 3.4 – Наиболее репрезентативные обобщения кластеров, полученные с помощью алгоритма ParGenFS. Приведены элементы кластеров со значением принадлежности  $u>0.15$

Интерпретация	Головные понятия и выбросы	Кол-во пробелов	Кол-во листьев в кластере
“Learning”	Гол. понятия: 1.1.1. – Machine learning theory 5.2. – Machine learning, выброс: 3.4.4.5. – Learning to rank	38	32
“Clustering”	Гол. понятие: 3.2.1.4. – Clustering выбросы: 1.1.1.3. – Unsupervised learning and clustering 2.1.5.8. – Cluster analysis 3.2.1.7.3 – Graph based conceptual clustering	0	17

Продолжение таблицы 3.4

	3.2.1.9.2. – Trajectory clustering 3.4.5.8. – Clustering and classification 5.2.1.2.1. – Cluster analysis 5.2.3.2.5 – Kernel-based clustering 5.2.4.3.1 – Spectral clustering		
Probabilistic representations	Гол. понятия: 5.2.1.2. – Unsupervised learning 5.2.3.5. – Learning in probabilistic graphical models Выбросы: 1.1.1.4.3. – Modelling 1.1.1.6. – Bayesian analysis 3.1.1.3.2. – Network data models 3.3.1.4. – Web log analysis 3.4.3.2. – Task models 5.2.3.1.3 – Model trees 5.2.3.13.1. – Deep belief networks 5.2.3.7.2. – Factor analysis	11	31
“Retrieval”	Гол. понятия: 3.1.4. – Query languages 3.4. – Information retrieval Выброс: 5.1.1.9. – Language resources	27	28

Полученная визуализация хорошо согласуется с общими представлениями о структуре Науки данных. Существует тесно связанный набор технологий, используемый для сбора и хранения необработанных данных. Далее, с помощью специализированных языков проводится обработка и агрегация данных. На основе обработанных данных строятся визуализации, которые являются наиболее простым и наглядным способом представления статистик, зависимостей и паттернов в данных. Кроме этого, обработанные данные могут использоваться для обучения без учителя и, в частности, для кластер-анализа, построения оценок плотности распределения данных и определения скрытых факторов, порождающих эти данные. Учитывая то, что количество неразмеченных данных с каждым днем неуклонно растет, можно заключить, что исследователи учитывают этот факт и в публикациях на тему извлечения и хранения неразмеченных больших данных зачастую поднимается тема их обработки.

### 3.1.4 Некоторые выводы

Найденные тематические кластеры показывают области наук о данных, которые часто исследуются в рамках одной и той же публикации, и могут

свидетельствовать о зарождении новой области исследований, находящейся на границе каждого из элементов кластера. Если же кластер расположен плотно (у него малое количество головных тем и выбросов), то он соответствует уже устоявшейся области исследований, которая хорошо отражена в использованной таксономии. Интерпретируя кластеры таким образом, можно заключить:

- кластер «Learning» соответствует уже устоявшейся теме исследований – методам обучения с учителем в задачах классификации, регрессии и ранжирования;

- «Clustering» отвечает одной из задач обучения без учителя – кластеризации. Наличие выбросов свидетельствует о том, что несмотря на то, что методы кластер-анализа используются в различных областях, кластеризация пока что не является отдельной единой областью наук о данных (возможно, скоро придет время выделить ее в отдельную вершину более высокого уровня);

- «Probabilistic representations» состоит из статистических методов представления данных (в том числе направленным графическим моделям и байесовским методам в статистике) и задач обучения без учителя. Это свидетельствует о том, что задачи обучения без учителя, постановка которых зачастую является эвристикой, постепенно получают математические обоснования и происходит соединение экспериментальной науки и теоретической;

- кластеры «Retrieval» и «Querying» соответствуют устоявшимся областям: извлечению данных и их обработке и агрегации;

- кластер «Computer vision representations» так соответствует устоявшейся теме исследований; можно заметить, что в выбросах этого кластера находятся вершины, отвечающие техникам нечетких и тензорных представлений данных;

- кластер «Structuring» отвечает сразу нескольким обширным темам: компьютерному зрению, теории баз данных, методам хранения данных и



различным представлениям данных. В том числе в выбросах присутствует темы, связанные с распознаванием речи и поиску в текстах. Все эти темы объединяет то, что объектом исследований в них являются различные методы работы с неструктурированными данными сложной структуры, для автоматического извлечения информации из которых требуются серьезные усилия и сложные алгоритмы. Можно надеяться на то, что в ближайшее время возникнет новая область исследований, в рамках которой будут разработаны общие подходы к решению проблемы обработки неструктурированных данных, которые будут использовать независимо от того, какого рода данные обрабатываются.

Если сравнивать данные результаты с выводами, полученными в ходе анализа первой коллекции, можно заключить следующее. Первые два из кластеров, найденных для второй коллекции, взаимно однозначно соответствуют кластерам L и C для первой коллекции.

Головные понятия и один выброс кластера L («Learning») в точности совпадает с обобщением, полученным для аналогичного кластера второй коллекции. Обобщение кластера C («Clustering») в данном расчете получилось более плотным: в нем всего 9 элементов (в предыдущем расчете получилось 16 элементов).

В то же время, третий кластер для первой коллекции, R ('Retrieval'), соответствует сразу пяти кластерам второй коллекции. Эти пять кластеров выглядят своеобразными «гранями» кластера R. Головная тема «Computer Vision», полученная для кластера R, в ходе анализа второй коллекции получила детализацию в виде «Structuring» и «Computer vision representations». Остальные кластеры являются уникальными для данного расчета.

Результаты работы были опубликованы в восьми статьях за 2019-2020 гг: [152] – [159].

## 3.2 Разработка и исследование метода альтернирующих дихотомий для кластеризации данных

### 3.2.1 Постановка задачи

В [160] - [165] был предложен достаточно общий подход к анализу сложности графов. Он состоит в построения семейства декомпозиций графов и в последующем вычислении показателей, характеризующих в том или ином смысле сложность данных графов и, следовательно, сложность исходных систем, моделируемых рассматриваемыми графами. Несмотря на полезные содержательные интерпретации, введенные показатели относятся к системе «в целом», а не к её отдельным элементам, взаимодействие которых и определяет функционирование рассматриваемых систем. Именно отсутствие информации о свойствах и взаимодействии отдельных элементов (а не только о системе в целом) является основным недостатком вышеупомянутого подхода. И преодоление этого недостатка является основной целью настоящей работы.

Остановимся подробнее на формальных определениях и постановках. Предполагается, что для любой пары элементов рассматриваемой системы определена некоторая степень близости между ними, во многом аналогичная традиционному расстоянию. Однако, в отличие от классического определения, нулевое расстояние между различными элементами возможно.

Упомянем два хорошо известных примера подобных систем. Для фондового рынка расстояние между двумя торгующимися акциями  $x$  и  $y$ , задаётся формулой  $d(x,y) = 1 - \rho(x,y)$ , где через  $\rho(x,y)$  обозначен коэффициент корреляции  $\rho(x,y)$  между этими акциями, подсчитанный по их стоимостям за некоторый период (за несколько дней, недель или месяцев). Значения  $d(x,y)$ , близкие к 0, означают высокую степень корреляции; значения  $d(x,y)$ , близкие к 2, означают противоположное поведение этих акций. Для голосующих органов (например, парламентов) за расстояние между двумя депутатами  $a$  и  $b$  можно принять обычное евклидово расстояние между двумя векторами  $v^a$  и  $v^b$ , соответствующими их  $n$  последовательным голосованиям за некоторый период, где

$$v_j = \begin{cases} 1, & \text{если депутат голосовал за } j\text{-ое предложение;} \\ -1, & \text{если депутат голосовал против } j\text{-го предложения;} \\ 0, & \text{в противном случае (воздержался или не участвовал).} \end{cases}$$

Понятно, что в обоих случаях нулевые расстояния являются хотя и не частыми, но достаточно естественными.

Одним из важнейших инструментов при исследовании разнообразных сложных систем является разбиение их элементов на группы в том или ином смысле похожих, т.е. решение задачи автоматической классификации (далее АК). Подробный содержательный обзор методов АК и многочисленные ссылки можно найти в монографии [166]. Для систем рассматриваемого типа можно выделить две основные группы методов. Одни основаны на построении графа соседства того или иного вида. При этом расстояния между элементами системы определяют моделирующий граф. Задаётся некоторое число  $d$  и ребро между вершинами проводится в том и только том случае, если расстояние между соответствующими элементами исходной системы не превосходит  $d$ . Далее, в рассматриваемом графе стараются найти клики или близкие к ним структуры. например, [167]. При использовании такого рода подхода удаётся найти несколько групп сильно связанных между собой элементов, но их доля мала по сравнению с общим числом элементов в системе.

Применение частотных методов дихотомии графа соседства [164],[165] может приводить к кластерам, содержащим достаточно длинные цепочки попарно близких элементов, так что расстояние между концами этих цепочек уже не является малым. Поэтому построенные таким образом кластеры не отвечают естественным требованиям к близости входящих в них элементов систем. А различные методы АК, не связанные с графами и основанные непосредственно на исходных расстояниях между элементами, не приводят во многих сложных случаях к достаточно ясным и хорошо интерпретируемым результатам.

Приведённые рассуждения инициировали поиск новых методов АК, применимых в рассматриваемых ситуациях. В предложенном методе АК сочетается как рассмотренный ранее частотный алгоритм дихотомии, так и чисто количественные соображения. Представляется, что метод позволяет найти разбиение всего множества элементов системы на непересекающиеся подмножества достаточно близких, так что построенные классификации определяются заданным уровнем близости элементов. Поскольку в предложенном методе чередуются частотный алгоритм дихотомии и численный анализ найденных кластеров, то его можно условно назвать альтернирующим (для краткости МАД – метод альтернирующих дихотомий).

В отчете приводится пример работы алгоритма для данных по фондовому рынку S&P-500.

### 3.2.2 Метод альтернирующих дихотомий

#### 3.2.2.1 Общая структура метода

Метод состоит в построении двоичного дерева, вершины которого соответствуют последовательно определяемым подмножествам множества вершин. Корень дерева соответствует множеству всех вершин рассматриваемого графа. Для каждого построенного подмножества вершин (обозначаемого через  $X$ ) определены два естественных показателя. Они таковы:

А. Среднее значение  $d(X)$  расстояний между всеми парами вершин для данного подмножества  $X$  (для одноэлементных подмножеств  $X$  по определению  $d(X) = 0$ ).

В. Число элементов  $N(X)$  в данном подмножестве  $X$ . Естественно, что это число известно.

Для использования метода должны быть известны:

1) Матрица  $D$  попарных расстояний между элементами исходного множества (для фондового рынка – матрица коэффициентов корреляции).

2) Максимально допустимое значение (обозначим его через  $d^*$ ) среднего расстояния для всех классов построенной классификации исходного множества вершин.

3) Максимально допустимое число элементов (обозначим его через  $N^*$ ) в каждом классе построенной классификации исходного множества вершин.

Естественно, что построенная классификация существенно зависит от параметров  $d^*$  и  $N^*$ . Понятно, что чем меньше  $d^*$  и  $N^*$ , тем больше должно быть число классов для удовлетворения вышеуказанных ограничений. Понятно также, что эти ограничения можно менять.

После этих предварительных замечаний переходим к непосредственному описанию МАД.

1) В качестве корня строящегося двоичного дерева берётся множество  $X^0$  всех вершин исходного графа. За его оценку по 1-му показателю принимается любое число, большее  $d^*$ , а по 2-му показателю – известное число  $N(X^0)$  всех вершин исходного графа.

2) Рассмотрим все листья уже построенного двоичного дерева. Пусть максимальная первая оценка (среднее попарное расстояние) для некоторого листа с множеством вершин  $X$  превосходит заданный уровень  $d^*$ . Если такого листа нет, переходим на шаг 4. Если такой лист есть, переходим на шаг 3, на котором выполняются следующие операции:

3.1) Строим граф соседства, исходя из множества вершин  $X$  и полной матрицы расстояний  $D$ .

3.2) Применяем к построенному графу частотный алгоритм дихотомии (ЧАД), кратко изложенный ниже в подразделе 2.2.

3.3) Результатом применения ЧАД является семейство пар множеств  $(X_1, X_2)$ . Если существуют пары, в которых оба множества  $X_1$  и  $X_2$  состоят из более чем одного элемента) переходим на шаг 3.4. В противном случае (это значит, что в каждой паре есть хотя бы одно одноэлементное множество) переходим на шаг 3.5.

3.4) Для каждой пары  $(X_1, X_2)$ , в которых оба множества не одноэлементны, подсчитаем оценки  $d(X_1)$  и  $d(X_2)$  (ещё раз напомним, что матрица  $D$  всех попарных расстояний известна, так что все вычисления осуществляются очень быстро). Переходим на шаг 3.6.

3.5) Для каждой пары  $(X_1, X_2)$ , в которых одно из множеств одноэлементно, подсчитаем оценки  $d(X_1)$  и  $d(X_2)$  (напомним, что для любого одноэлементного множества  $X$   $d(X) = 0$ ).

3.6) Выберем ту пару  $(X_1, X_2)$ , для которой сумма  $d(X_1) + d(X_2)$  минимальна.

3.7) Найденные подмножества вершин  $X_1$  и  $X_2$  добавляются к строящемуся двоичному дереву как сыновья вершины  $X$ . Их первые оценки уже подсчитаны, а вторые (число элементов) известны, поскольку известны сами множества.

3.8) Возвращаемся к шагу 2.

4) Рассмотрим все листья уже построенного двоичного дерева. Пусть максимальная вторая оценка (число элементов в множестве) для некоторого листа с множеством вершин  $X$  превосходит заданный уровень  $N^*$ . Если такого листа нет, переходим на шаг 6. Если такой лист есть, переходим на шаг 5, на котором применительно к множеству вершин  $X$  выполняются все операции шагов 3.1 – 3.8.

5) Окончание работы алгоритма. Все листья образуют классификацию исходного множества вершин, в которой удовлетворяются требования как по среднему попарному расстоянию, так и по числу вершин в каждой части.

Комментарии к МАД.

1) Сходимость метода связана со следующими простыми обстоятельствами. Если хотя бы одно из двух условий не удовлетворяется, то осуществляется деление хотя бы одного множества. Поскольку операций объединения в МАД нет, то после конечного числа шагов гарантируется выполнение обоих условий – хотя бы тривиальным образом на одноэлементных подмножествах. Однако в экспериментах с реальными данными таких примеров не встретилось, т.е. остановка при достижении заданных уровней происходила раньше.

2) При наличии хотя бы одной пары с обоими не одноэлементными множествами, те пары, в которых одно из множеств одноэлементно, далее не рассматриваются, несмотря на то, что соответствующие суммы для них могут

быть меньшими. Этим устраняются часто встречающиеся при различных методах классификации сильно неоднородные варианты.

3) Вместо выбора (на шаге 3.6) пары  $(X_1, X_2)$  с минимальной суммой  $d(X_1) + d(X_2)$  можно выбирать пару  $(X_1, X_2)$ , у которой  $\max\{d(X_1), d(X_2)\}$  минимален.

4) Поскольку результат работы ЧАД из-за многократного использования в нём случайного датчика неоднозначен, то неоднозначны и результаты выполнения МАД. Однако в любом случае происходит разбиение на части, удовлетворяющие заданным условиям как по числу элементов, так и по средним попарным расстояниям.

5) Подробный пример работы МАД приводится ниже.

#### 3.2.2.2 Частотный алгоритм дихотомии

В настоящем подпункте кратко изложим алгоритм построения семейства дихотомий произвольного неориентированного простого графа, опубликованный в [165]. Поскольку графы других типов не рассматриваются, будем далее пользоваться термином «граф» без дополнительных уточнений.

Дадим описание общей структуры ЧАД. Предполагается, что исходный граф и число  $M$  строящихся дихотомий заданы.

1) На предварительном этапе каждому ребру исходного графа сопоставляется случайное целое число от 1 до 5 включительно. Далее числа, сопоставляемые рёбрам графа, будут называться частотой соответствующего ребра. Обозначим через  $F_{\max}$  максимальную частоту в рёбрах графа.

2) 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 (пункт 1)

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-й 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 на шаге 4, 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.

Некоторые пояснения и иллюстрации к предложенному алгоритму приведены в статье [165].

Алгоритм останавливается, когда построено заданное число  $M$  дихотомий. Многие из построенных дихотомий могут совпадать.



### 3.2.3 Применение МАД для анализа фондового рынка

Для фондового рынка привычнее говорить о коэффициенте корреляции между последовательными курсами для двух акций, чем о расстоянии между ними, что совершенно не влияет на описанный выше МАК. Просто вместо минимизации среднего расстояния можно говорить о максимизации среднего коэффициента корреляции и т.д. Напомним, что для фондового рынка за расстояние между двумя элементами  $x$  и  $y$  принимается величина  $d(x,y) = 1 - \rho(x,y)$ .

Для примера рассмотрим рынок S&P-500 (500 крупнейших компаний в США). Рассматривается фиксированная дата  $x$ . Коэффициенты корреляции определялись по ценам на 492 акции на момент закрытия в течении 16-и последовательных дней, из которых день  $x$  – последний день данного 16-дневного периода. Симметричная матрица корреляций  $R$  предполагается известной. Она легко вычисляется по курсам акций за данный период. Подробное описание ситуации приведено в статье [165] и в прошлогоднем отчёте. Для демонстрации МАД достаточно знать матрицу корреляций  $R$  размера  $492 \times 492$ . Зададим также нижнее ограничение на средние коэффициенты корреляции в финальных листьях строящегося двоичного дерева – 0.60 и верхнее ограничение на число элементов в финальных листьях – 164.

В соответствии с описанием МАД на шаге 1 корнем строящегося дерева является множество всех вершин соответствующего графа. Этот корень показан на рисунке 3.8.

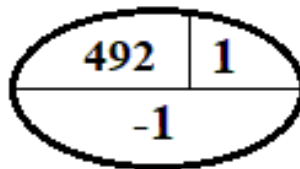


Рисунок 3.8 – Корень двоичного дерева

Вершина разделена на три поля. В верхнем левом поле приводится число элементов в соответствующем множестве (в данном случае их начальное число 492), в нижнем поле – число -1, поскольку это число меньше заданного уровня 0.60. Заполнение правого верхнего поля объясняется ниже. Переходим на шаг 2.

Поскольку 1-ое условие не удовлетворяется ( $-1 < 0.60$ ), то в соответствии с описанием на шаге 2 МАД (здесь и далее для сравнения по 1-му показателю надо рассматривать не максимальное, а минимальное значение) в единственном листе (он совпадает с корнем) имеем значение 1-го показателя  $d(X^0) = -1 < d^*$ . Поэтому записываем в верхнюю правую часть число 1, означающее, что переход на шаг 3 делается 1-й раз.

В результате выполнения всех операций шага 3 (начиная с построения графа соседства и применения к нему ЧАД) получаем следующее двоичное дерево, показанное на рисунке 3.9.

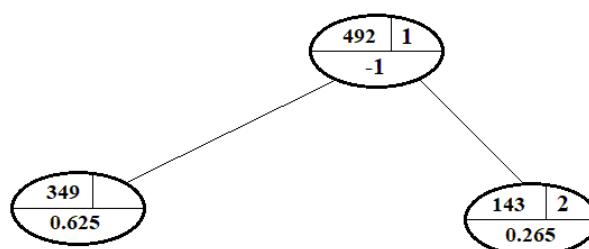


Рисунок 3.9 – Двоичное дерево после 1-го деления

В новом дереве имеется два листа – с 349 и 143 вершинами. Поскольку минимальным является 1-й показатель у листа со 143 вершинами и он меньше заданного ограничения 0.6, то в соответствии с описанием на шаге 2 МАД для следующего деления выбираем этот лист. Записываем в правое верхнее поле выбранного листа число 2, означающее, что именно из этого листа делается 2-е деление.

В результате выполнения всех операций шага 3 получаем следующее двоичное дерево, показанное на рисунке 3.10.

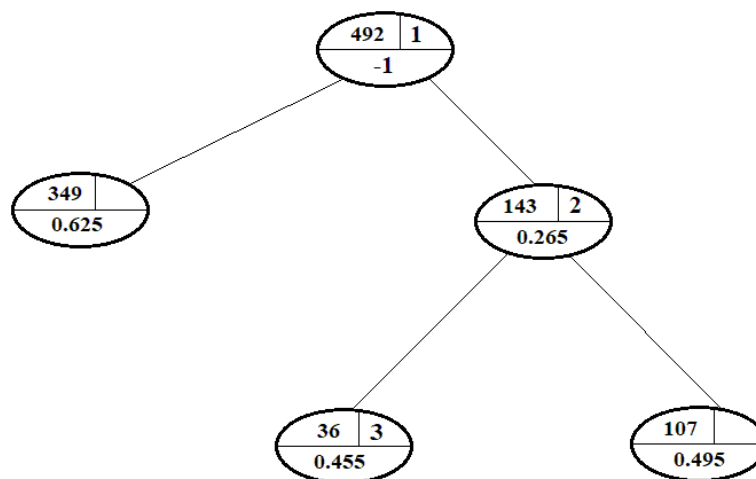


Рисунок 3.10 – Двоичное дерево после 2-го деления

В новом дереве имеется три листа – с 349, 36 и 107 вершинами. Поскольку минимальным является 1-й показатель у листа с 36 вершинами и он меньше заданного ограничения 0.6, то в соответствии с описанием на шаге 2 МАД для следующего деления выбираем этот лист. Записываем в правое верхнее поле выбранного листа число 3, означающее, что именно из этого листа делается 3-е деление. Обратим внимание, что в остальных двух листьях в правом верхнем поле ничего не записано. Заметим также, что после внесения записи в это поле в результате деления данной вершины у неё появляется два сына и данная вершина перестаёт быть листом в строящемся двоичном дереве.

В результате выполнения всех операций шага 3 получаем следующее двоичное дерево, показанное на рисунке 3.11.

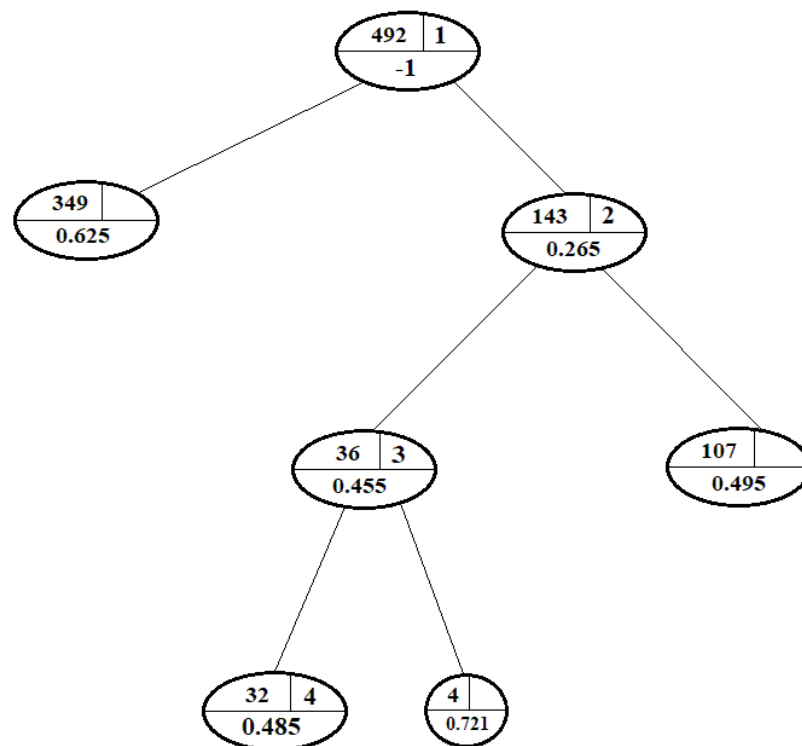


Рисунок 3.11 – Двоичное дерево после 3-го деления

В новом дереве имеется четыре листа – с 349, 32, 4 и 107 вершинами. Заметим сразу, что лист с 4 вершинами изображён по-другому – не овалом, а кружком. Суть дела в том, что для этого листа условия на оба показателя выполнены:  $0.721 > 0.6$  и  $4 < 164$ . В соответствии с МАД этот лист при данных ограничениях  $d^* = 0.6$  и  $N^* = 164$  никогда не будет делиться. Рассмотрим оставшиеся три листа. Поскольку минимальным является 1-й показатель (0.485) у листа с 32 вершинами и он меньше заданного ограничения 0.6, то в соответствии с описанием на шаге 2 МАД для следующего деления выбирается этот лист. Записываем в правое верхнее поле выбранного листа число 4, означающее, что именно из этого листа делается 4 деление. Обратим внимание, что в остальных двух листьях (с 349 и 17 вершинами в правом верхнем поле) ничего не записано. Как и ранее, после внесения записи в это поле в результате деления данной вершины у неё появляется два сына, и данная вершина перестаёт быть листом в строящемся двоичном дереве.

В результате выполнения всех операций шага 3 получаем следующее двоичное дерево, показанное на рисунке 3.12.

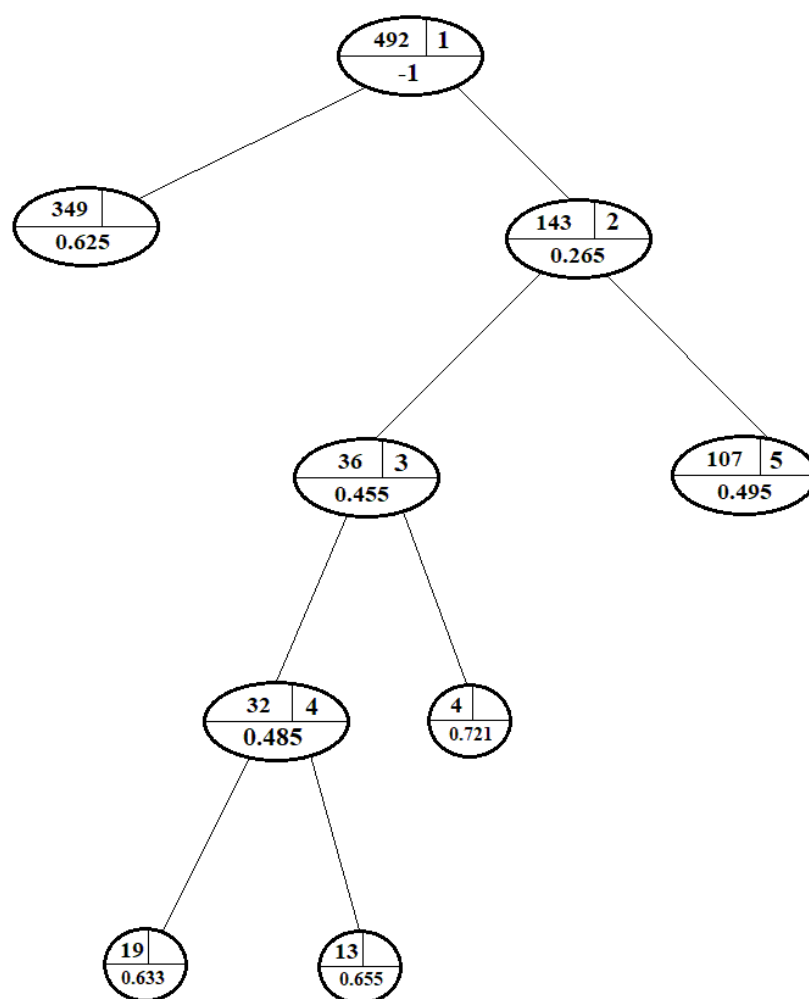


Рисунок 3.12 – Двоичное дерево после 4-го деления

В новом дереве имеется пять листьев – с 349, 19, 13, 4 и 107 вершинами. Листья с 19, 13 и 4 вершинами не могут быть выбраны из тех же соображений, по которым в предыдущем дереве не мог быть выбран лист с 4 вершинами. Остаются два листа – с 349 и со 197 вершинами. Поскольку минимальным является 1-й показатель (0.495) у листа со 107-ю вершинами и он меньше заданного ограничения 0.6, то в соответствии с описанием на шаге 2 МАД для следующего деления выбирается этот лист. Записываем в правое верхнее поле выбранного листа число 5, означающее, что именно из этого листа делается 5-ое деление. Как и ранее, после внесения записи в это поле в результате деления данной вершины у неё появляется два сына, и данная вершина перестаёт быть листом в строящемся двоичном дереве.

В результате выполнения всех операций шага 3 получаем следующее двоичное дерево, показанное на рисунок 3.13.

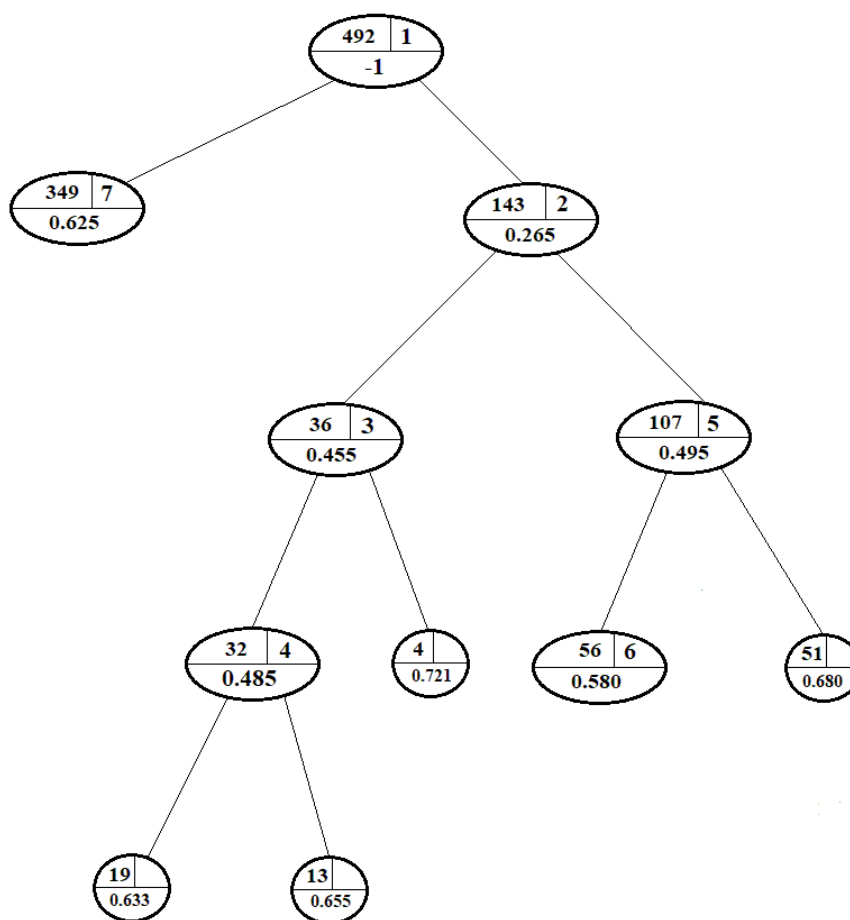


Рисунок 3.13 – Двоичное дерево после 5-го деления

Аналогично предыдущим шагам, легко видеть, что следующим делится лист с 56-ю вершинами и 1-м показателем 0.580.

В результате выполнения всех операций шага 3 получаем следующее двоичное дерево, показанное на рисунок 3.14.

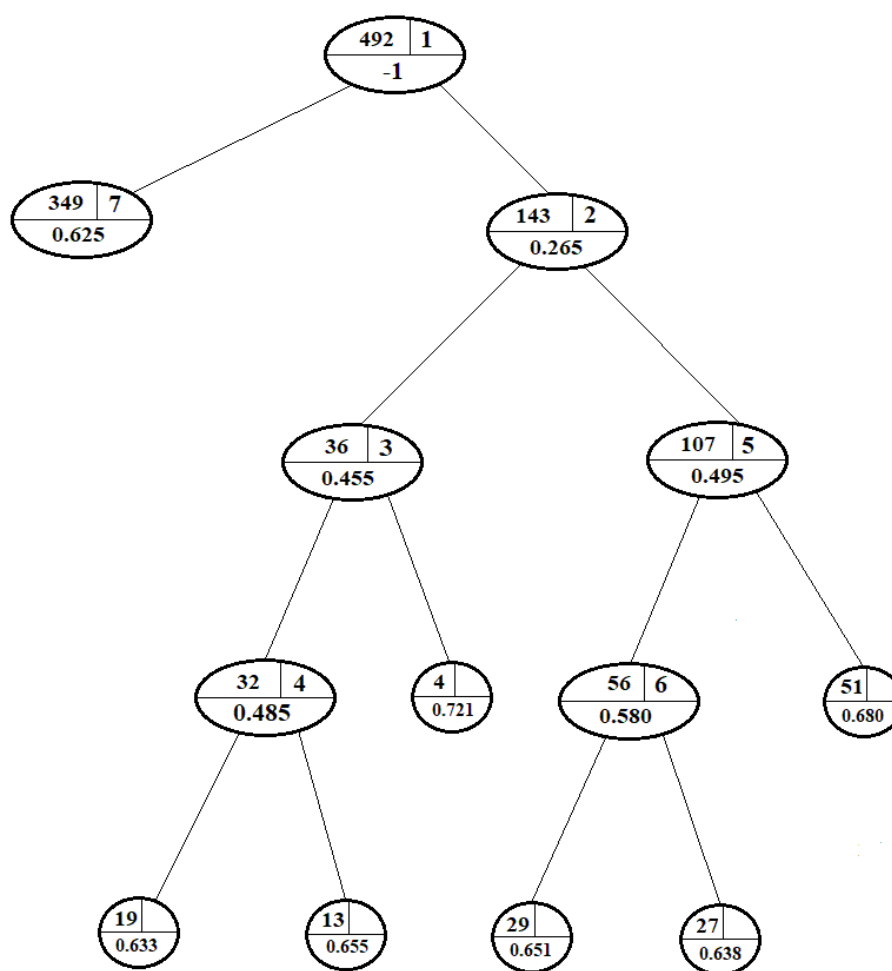


Рисунок 3.14 – Двоичное дерево после 6-го деления

Ситуация после 6-го деления изменилась. Во всех 7 листьях последнего двоичного дерева 1-й показатель  $d$  – среднее значение коэффициента корреляции у соответствующих подмножеств акций – превосходит заданное ограничение 0.6. В соответствии с МАД в качестве следующего деления выбирается деление листа с 349 вершинами, поскольку число вершин превосходит заданный верхний уровень 164. В результате деления (как и во всех предыдущих случаях, основанного на ЧАД) получаем двоичное дерево, показанное на рисунке 3.15.

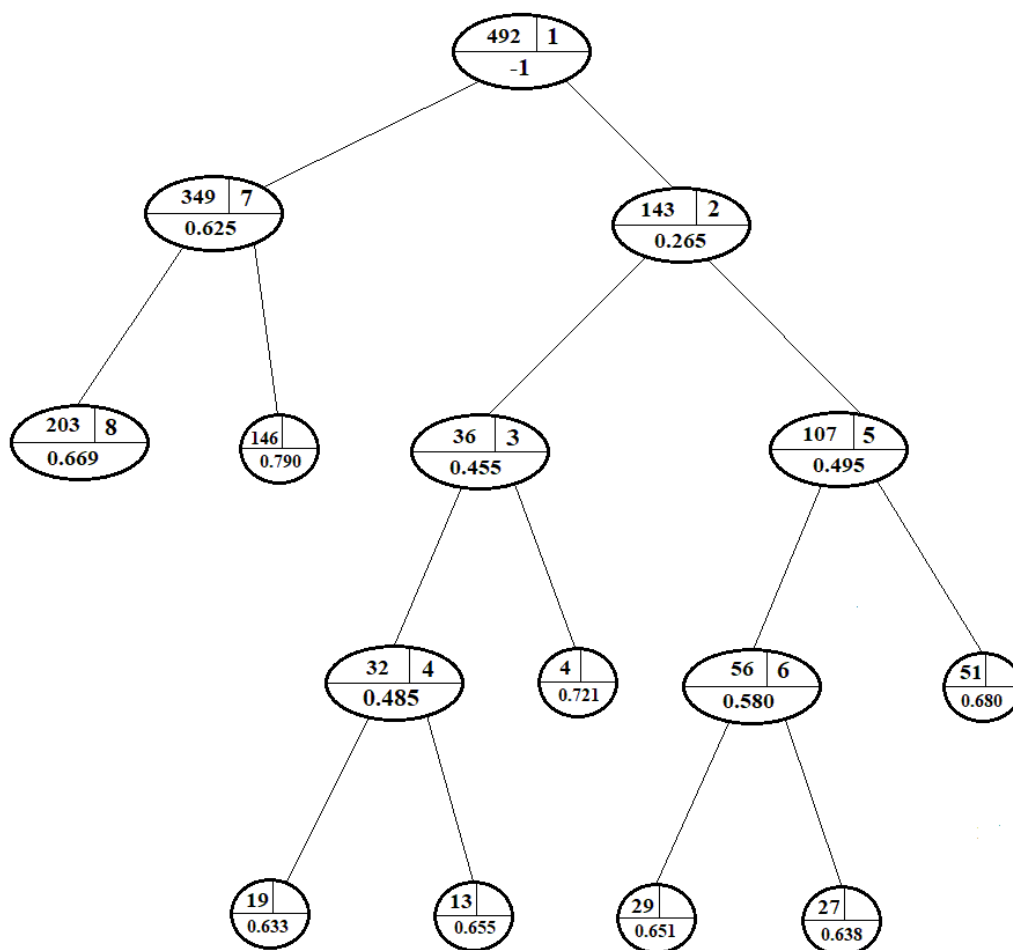


Рисунок 3.15 – Двоичное дерево после 7-го деления

Как и в предыдущем дереве, имеется лишь один доступный для деления лист с 203 вершинами. После его деления получаем двоичное дерево, показанное на рисунке 3.16.

Нетрудно увидеть, что в последнем дереве нет листьев, которые можно было бы выбрать для последующего деления. В соответствии с шагом 6 МАД происходит окончание работы алгоритма. Все листья образуют разбиение исходного множества вершин, состоящее из 9 частей, в котором удовлетворяются требования как по среднему попарному коэффициенту корреляции, так и по числу вершин – то и другое в каждой части. Естественно, что при увеличении уровня  $d^*$  и уменьшении уровня  $N^*$  число частей может увеличиться.



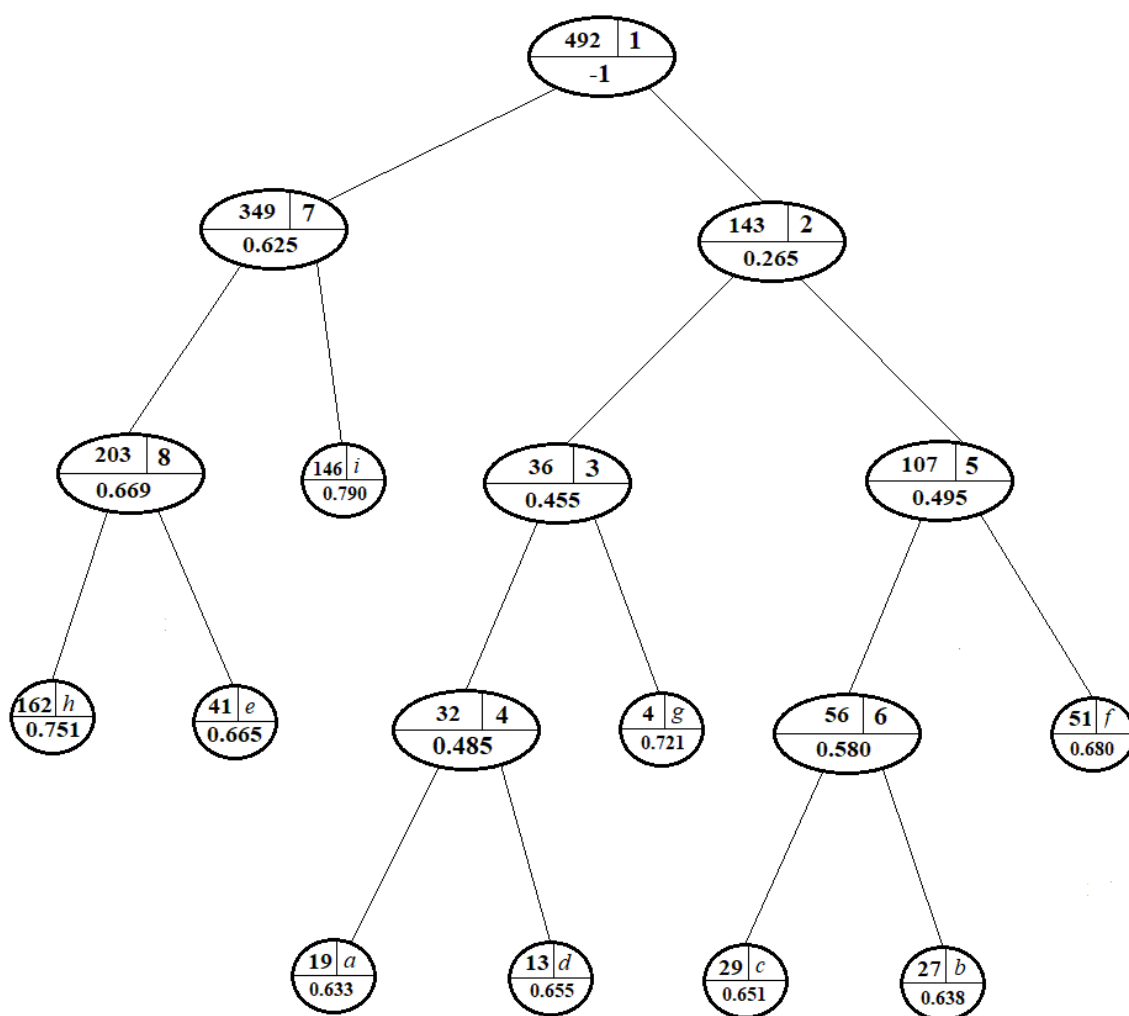


Рисунок 3.16 – Окончательное двоичное дерево после 8-го деления

На рисунке 3.16 видны все 9 листьев, образующих разбиение исходного множества вершин. Латинские буквы на листьях соответствуют нумерации в порядке возрастания средних коэффициентов корреляции, которые изменяются от 0.633 до 0.790. Однако для полноты картины желательно указать средние коэффициенты корреляции между найденными кластерами. Они легко вычисляются, поскольку известны все кластеры и исходная матрица коэффициентов корреляции  $R$  для всех пар акций, на основе которой и осуществлены все построения в соответствии с МАД. Результаты вычисления средних межкластерных корреляций представлены в Таблице 3.5.

Таблица 3.5 – Средние межкластерные корреляции

	<i>b</i>	<i>c</i>	<i>d</i>	<i>E</i>	<i>f</i>	<i>G</i>	<i>h</i>	<i>i</i>
<i>a</i>	0.200	0.175	0.350	- 0.253	-0.297	0.353	-0.427	-0.618
<i>b</i>		0.518	-0.134	0.093	0.429	0.205	-0.344	0.003
<i>c</i>			-0.289	-0.262	0.308	0.197	-0.538	-0.110
<i>d</i>				0.090	-0.300	0.289	0.195	-0.229
<i>e</i>					0.374	-0.042	0.507	0.487
<i>f</i>						-0.001	0.108	0.538
<i>g</i>							-0.151	-0.236
<i>h</i>								0.547

Таблица 3.5 демонстрирует отсутствие сколько-нибудь заметных корреляций между разными кластерами. Среди 36 чисел есть всего лишь 6, по модулю бóльших, чем 0.5, из которых 4 положительных и 2 отрицательных. Можно сказать, что все построенные 9 кластеров хорошо различимы.

### 3.2.4 Некоторые выводы

Приведенные выше соображения являются лишь первой попыткой изложения нового метода автоматической классификации, применимой к ситуациям, в которых тем или иным способом заданы значения похожести/непохожести для всех пар элементов. Конечно, в рамках предложенного подхода следует провести ещё многие исследования. Назовём несколько основных направлений.

1) Выше упоминалось о возможности построения нескольких классификаций. Желательно понять, как они соотносятся между собой. В частности, возможны две ситуации – при ослаблении порогов и, следовательно, увеличении числа кластеров, построенные классификации становятся всё более близкими или же имеется несколько различных классификаций, на которые похожи все остальные.

2) Необходимо расширить список приложений – уже потому, что ситуации, когда требуется решить задачу классификации при наличии известной матрицы похожести/непохожести, является распространёнными и разнообразными.

3) Следует попытаться обобщить предложенный подход на ситуации, в которых похожесть/непохожесть элементов определяется не в количественных, а в качественных терминах.

### **3.3 Исследование государственной состоятельности в странах мира методами кластерного анализа**

#### **3.3.1 Обзор литературы**

В настоящее время возрастает интерес к изучению государственной состоятельности со стороны отечественных и зарубежных ученых. К примеру, запрос «State Capacity» в ресурсе Google Academy выдает 3.930.000 результатов (143.000 – только с начала 2019 года). Несмотря на большое количество работ, посвященных данной теме, возникают множество проблем с операционализацией самого понятия «государственная состоятельность». В [168] предлагается следующее определение: «способность государства выбирать и эффективно реализовывать собственные решения, в том числе касающиеся внутренней и внешней политики».

В связи со сложностью определения самого понятия «государственной состоятельности» (хотя и существует множество подходов [169],[170]) возникают и сложности в определении системы показателей, позволяющей дать количественную оценку исследуемых стран. С этой целью, в работе составлена базовая система показателей, проведен корреляционный анализ и исключены показатели, имеющие высоко коррелированные значения. Итогом является определение 5 базовых показателей: качество государственных институтов; доля налоговых поступлений в ВВП; доля военных расходов в ВВП; обратное значение суммарного уровня жертв конфликтов и убийств; доля военного персонала в общей численности населения. Данная система составлена для определения групп стран, схожих (с использованием определенных мер близости) и изучения динамических траекторий развития каждой из них. Для решения данной задачи возможно использование нескольких основных подходов. К наиболее популярным можно отнести:

– использование методов классификации данных. Применение, к примеру, нейронных сетей к исходному набору стран позволяет построить модель соотнесения исследуемых объектов к различным классам. Несмотря на многие преимущества подобных методов, для их реализации требуется сформировать обучающую выборку объектов (что в рассматриваемой задаче весьма непросто). В общем случае, для использования методов классификации данных, нам требуется заранее знать количество искомых классов и их типичных представителей;

– составление единого агрегированного рейтинга с использованием линейной свертки (и аналогов). В общем случае, соотнесение единого числового, дающего возможность составления обобщенного рейтинга стран в зависимости от значения данного показателя является весьма популярной задачей. Несмотря на простоту использования данного подхода, существенным его недостатком (после сложности в определении исходной системы показателей) является выбор весов каждого критерия. Даже если предположить, что веса можно определить (или сделать равнозначными), при использовании данного подхода для поиска объектов со схожими структурами возникает следующая проблема: низкие значения одних параметров могут быть компенсированы высокими значениями других;

– использование методов индивидуального и коллективного выбора. Поскольку в задаче исследуются многомерные числовые данные, то можно провести ранжирование стран по каждому отдельному показателю, после чего использовать различные модели агрегирования (к примеру, правило Борда, Хара, Кумбса, Коупленда, Нансона и др.). При необходимости отсутствия возможности компенсирования низких значений одних параметров высокими значениями других, возможно использование порогового агрегирования. Однако, при наличии одинаковых числовых значений для разных показателей исследуемые объекты будут соотнесены к единой группе;

– использование методов кластерного анализа. Применение различных существующих методов кластеризации к исходному набору данных позволяет

выявить группы объектов со схожими (по некоторой мере близости) значениями параметров. Из [171]: «Под кластером обычно понимается часть данных (в типичном случае – подмножество объектов или подмножество переменных, или подмножество объектов, характеризуемых подмножеством переменных), которая выделяется из остальной части наличием некоторой однородности элементов»;

– использование методов анализа паттернов. Подобные методы могут служить возможным дополнением к методам кластеризации данных для сравнения и выявления объектов, соотнесенных к единой группе на основе различных метрик. Особенностью данного подхода является объединение в единую группу объектов, существенно отличающихся по абсолютным значениям показателей, но имеющих схожие структуры данных. Из [172]: «Паттерн – комбинация определенных качественно похожих признаков».

В данной части отчета приведено сопоставление результатов порядково-инвариантной паттерн-кластеризации и методов иерархической кластеризации (с использованием различных мер близости) для формирования динамических траекторий развития стран. Основные результаты разбиения представлены в [219].

### 3.3.2 Общая постановка задачи исследования

Одной из основных целей данного исследования – определить траекторию развития государственной состоятельности в странах мира на основе выбранной системы показателей и сопоставление результатов, полученных 2 разными методами: кластерного анализа и анализа паттернов. Опишем задачу в общем виде.

В качестве исходных данных взято множество объектов типа «стран\_год»  $W:|W|=482$ . Опорными точками выбраны данные за 1996, 2005 и 2015 гг. Всего исследуется 150 стран в 1996, и 166 – в 2005 и 2015. Для удобства, объекты заданы в векторном виде  $w_i = (w_{i1}, w_{i2}, w_{i3}, w_{i4}, w_{i5})$ ,

где

$w_{i1}$  – качество государственных институтов  $i$ -ого объекта;

$w_{i2}$  – доля налоговых поступлений в ВВП  $i$ -ого объекта;

$w_{i3}$  – доля военных расходов в ВВП  $i$ -ого объекта;

$w_{i4}$  – обратное значение суммарного уровня жертв конфликтов и убийств  $i$ -ого объекта;

$w_{i5}$  – доля военного персонала в общей численности населения  $i$ -ого объекта.

Задачей анализа является, с одной стороны, выявление стран, схожих по уровню государственной состоятельности (при это для определения меры сходства используются различные, заранее заданные метрики), с другой – выявление траекторий развития государственной состоятельности в исследуемых странах (т.е. смены стратегии развития относительно выбранной системы показателей).

Для выполнения поставленных задач, проводится тщательный анализ данных, и исключаются выбросы. Далее, проводится корреляционный анализ (для анализа паттернов рекомендуется использовать данные с относительно невысокими коэффициентами корреляции) и последующая нормировка. В настоящей работе использовалось так называемое линейное масштабирование согласно формуле

$$w_{ij}^* = \frac{w_{ij} - w_{\min\_j}}{w_{\max\_j} - w_{\min\_j}}, \quad (3.7)$$

где  $w_{ij}^*$  – полученное с использованием линейного масштабирования нормированное значение  $j$ -ого параметра  $i$ -ого объекта.

Далее, с использование определенных мер близости (описанных ниже, различных для методов анализа паттернов и кластеризации, примененных в данной работе) определяются группы стран со схожей структурой государственной состоятельности. Поскольку не существует общепризнанной системы показателей, характеризующей государственную состоятельность (как и общего количественного метода для сопоставления уровня в различных

странах мира), в работе сопоставляются результаты, полученные на основе методов анализа паттернов и кластерного анализа для выявления относительно «устойчивых» групп.

Последним шагом является исследование траектории развития уровня государственной состоятельности в странах мира. Следует отметить, что целью данной работы не является построение общего рейтинга стран по уровню государственной состоятельности или сравнение полученных групп на предмет «лучше/хуже», а лишь выявление групп со схожей структурой (по выбранной системе показателей, характеризующей государственную состоятельность) и выявление смены структуры государственной состоятельности в 1996-2015гг.

### 3.3.3 Оценивание государственной состоятельности методами анализа паттернов

В настоящее время, множество работ посвящены методам анализа паттернов. Несмотря на наличие некоторых работ, демонстрирующих возможность использования данного метода при анализе электорального поведения ([174]), в политической науке данный метод использовался не очень часто.

В общем случае, под паттерном понимается «комбинация определенных качественно похожих признаков» ([175]). Таким образом, в основе данных методов – объединение именно качественно (а не количественно) близких объектов. К рассматриваемым в настоящей работе методам предъявлены также некоторые требования: устойчивость результатов к выбору последовательности входных данных; отсутствие необходимости задания реализующему алгоритму количества конечных паттернов; относительно высокая скорость работы алгоритма; отсутствие необходимости задания состава (или типичных представителей) конечных паттернов.

Как описано выше, исследуется 462 объекта типа «страна\_год», по 5 параметрам, характеризующим государственную состоятельность. Методы анализа паттернов, как правило, предполагают использование системы парал-

лельных координат [176] для визуализации исследуемых объектов. Данная система содержит (в настоящем исследовании) 5 осей (по количеству параметров), каждая из которых описывает отдельно взятый параметр. Поскольку каждому объекту типа «страна\_год» поставлен в соответствии 5-мерный вектор  $w_i = (w_{i1}, w_{i2}, w_{i3}, w_{i4}, w_{i5})$ , можно построить через точки  $w_{i1}, w_{i2}, w_{i3}, w_{i4}, w_{i5}$  некоторую кривую. Другими словами, функцию  $g^w: R \rightarrow R$ , такую, что  $g^w(i) = w_{i1}$  для всех  $i=1,2,\dots,5$ .

После визуализации встает вопрос о выборе метрики разбиения для получения результатов. В анализе паттернов есть важная особенность: результат не должен зависеть от абсолютных значений выбранных параметров. К примеру, если исследуются гипотетические объекты  $w_1 = (1, 2, 3, 4, 5)$  и  $w_2 = (10, 20, 30, 40, 50)$ , то (в зависимости от постановки задачи) большая часть методов кластеризации соотнесет данные объекты в разные группы (кластеры). Однако, если посмотреть на значения параметров, очевидно, что объекты  $w_1$  и  $w_2$  имеют схожую структуру (значения параметров второго объекта есть значения параметров первого объекта, умноженные на 10). Методы анализа паттернов должны объединять подобные объекты в единую группу.

В настоящей работе используется порядково-фиксированная и порядково-инвариантная паттерн-кластеризации, подробно описанные в [175]. В основе данных методов лежит парное сравнение всех исследуемых параметров и формирование кодировок объектов для объединения в группы. В качестве меры близости выступает расстояние Хемминга между кодировками объектов. Однако имеется важное дополнение: корректировка результатов осуществляется с использованием центроидов (методология описана в [175]). После разбиения объектов по группам проводится динамический анализ. С этой целью, формируются траектории развития исследуемого объекта по выбранным показателям, и производится поиск объектов с устойчивыми/неустойчивыми траекториями.



### 3.3.4 Оценивание государственной состоятельности с помощью иерархической кластеризации

Для сопоставления результатов, полученных на основе порядково-инвариантной паттерн-кластеризации, в работе используются методы кластерного анализа, успешно примененные в ряде областей (в т.ч. и в политологии). Методам кластерного анализа посвящены множество работ [177], [178], в связи с чем не будем их подробно описывать в настоящей работе. Однако в нескольких словах приведем методологию, используемую здесь.

Для начала, требуется выбрать конкретную метрику, и соответственно метода кластеризации для решения поставленных в работе задач. Разумеется, вопрос о выборе наиболее подходящего метода задавался уже давно. Одним из классических примеров является гипотетический пример о разбиении некоторого набора точек, представленных на рисунке 3.17 [179]. На данном примере видно, что использование различных методов кластеризации к одному набору данных может приводить к весьма различным результатам. В таком случае, вполне обоснованным кажется вопрос о целесообразности использования тех или иных методов и какой метод использовать в каждом конкретном случае. Более того, при сопоставлении результатов важно понять, какой из них можно использовать при решении поставленной задачи.

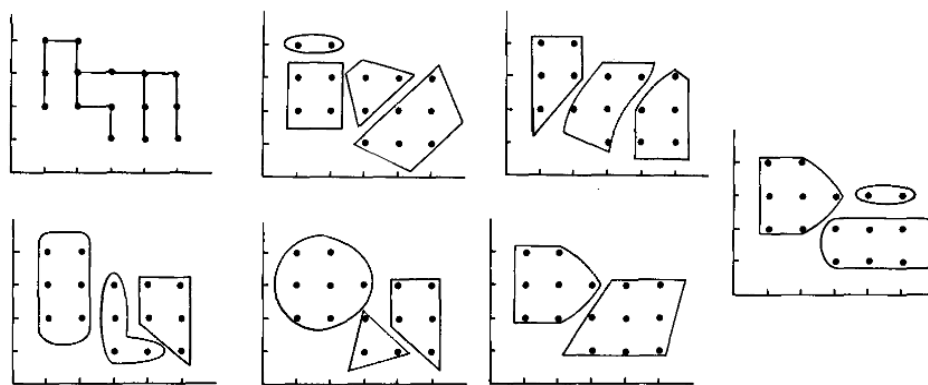


Рисунок 3.17 – Пример использования различных методов кластеризации на одном наборе данных. Рисунок модифицирован из [179]

Разумеется, ответить на данный вопрос весьма непросто. В качестве общей рекомендации, можно использовать результат, который имеет хорошо обоснованную интерпретацию.

В настоящей работе было выбрано использование иерархической кластеризации. С этой целью, на базе имеющихся данных была построена дендрограмма (рисунок 3.18).

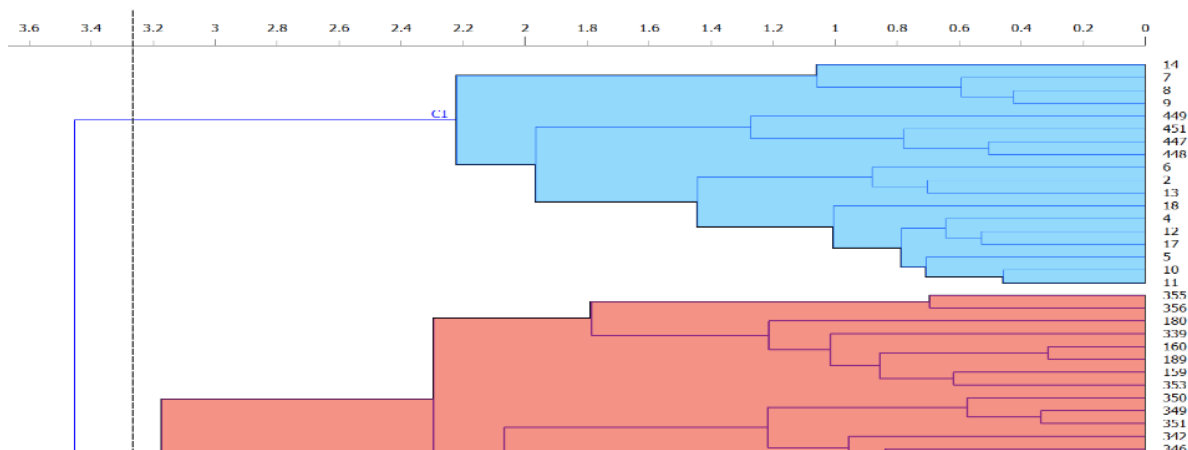


Рисунок 3.18 – Пример части построения дендрограммы для исследуемых объектов

В качестве меры близости при использовании иерархической кластеризации в настоящей работе выступает расстояние Евклида (однако, проводилось исследование с использованием других метрик). В результате исходное множество объектов было разбито на основных 9 кластеров, которые в дальнейшем были сопоставлены с группами, полученными с использованием порядково-инвариантной паттерн-кластеризации (с корректировкой результатов на основе центроидов полученных порядково-инвариантных паттерн-кластеров). Данное сопоставление сделано для получения некоторых «устойчивых» результатов.

### 3.3.5 Сопоставление результатов порядково-инвариантной паттерн-кластеризации и иерархического кластерного анализа

Приведем некоторые группы объектов, полученных на основе методов анализа паттернов и кластеризации. Несмотря на разность примененных в работе подходов, при корректировке результатов порядково-инвариантной паттерн-кластеризации на центры каждой группы, весьма разные методы дали очень схожие результаты (рисунок 3.19).

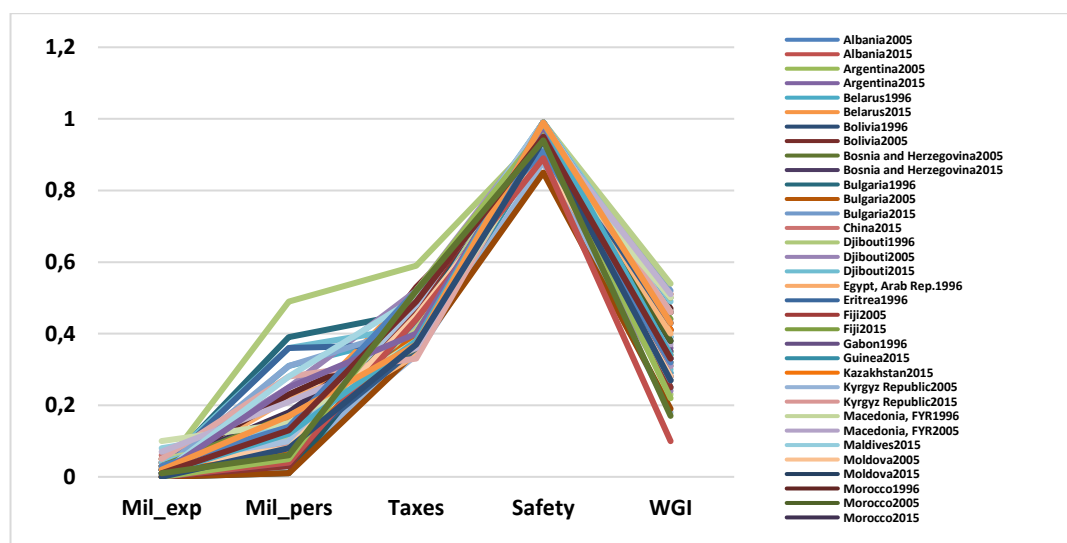


Рисунок 3.19 – Группа «1». Пример объединения стран по уровню государственной состоятельности

Примененные в работе методы отнесли данные объекты в единую группу. Визуально можно отметить отличительные особенности: относительно невысоки значения показателей «Mil\_exp» и «Mil\_pers», относительно средние значения «Taxes» и «WGI», а также относительно высокие значения «Safety». Данная группа характеризует 34 страны (63 объекта).

Продемонстрируем ещё один пример. На рисунке 3.20 видно объединение объектов на основе 2 различных методов. Характерным являются крайне высокие значения показателей «WGI» и «Safety», высокие и относительно высокие значения «Taxes», а также низкие и относительно средние значения «Mil\_exp» и «Mil\_pers». В данную группу вошли 17 стран (всего 50 объектов), все из которых (за исключением Ирландии) не меняют принадлежность к данной группе за исследуемый период: Дания, Австрия, Люксембург, Франция,

Норвегия, Бельгия, Швеция, Австралия, Новая Зеландия, Канада, Германия, Ирландия, Швейцария, Нидерланды, Исландия, Финляндия и Великобритания.

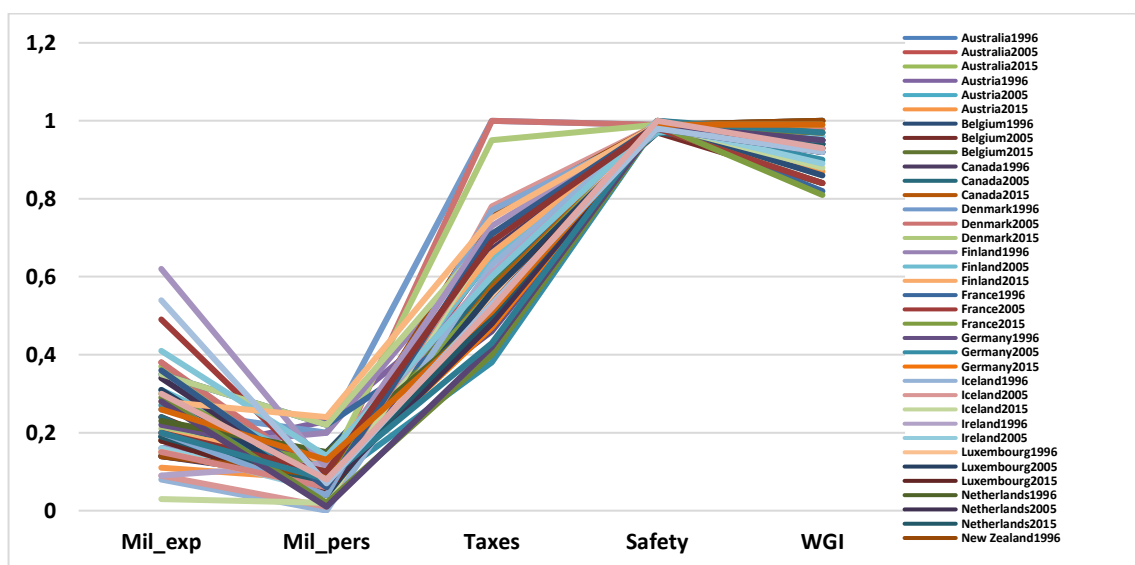


Рисунок 3.20 – Другой пример объединение объектов в группы на основе 2 методов

Подробное описание полученного разбиения на исследуемых стран на группы представлено в [173].

### 3.3.6 Динамический анализа траекторий развития стран

Итогом данного исследования является изучение траекторий развития государственной состоятельности в странах мира. С этой целью, на основе полученных ранее результатов, полученных с использованием иерархической кластеризации и порядково-инвариантной паттерн-кластеризации. Таким образом, исследованы группы типа «а-а-а» (не меняющую свою траекторию развития), «а-б-б» (меняющие единожды), «а-б-а» (меняют, но возвращаются к стратегии 1996 года), «а-б-с» (меняющие стратегию в каждый рассматриваемый период времени. Пример полученных результатов приведен на рисунке 3.21.

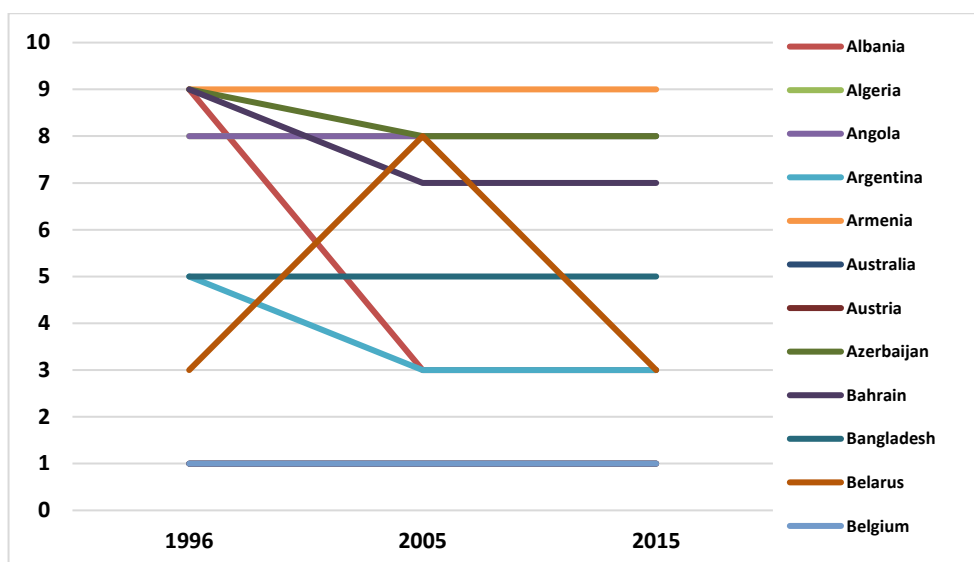


Рисунок 3.21 – Динамические траектории развития государственной состоятельности в некоторых странах

Из рисунка 3.21 видно, что такие страны, как Бельгия и Армения не меняли принадлежности к одной группе за весь исследуемый период (по выбранной системе показателей), Белоруссия – меняет один раз. Подобный анализ проведен для 166 стран мира.

### 3.3.7 Некоторые выводы

В данной части отчета приведены результаты исследования государственной состоятельности в странах мира в период 1996-2015гг. Определена базовая система показателей, предложено сопоставление двух различных методов (иерархической кластеризации и порядково-инвариантной паттерн-кластеризации) для получения устойчивых результатов. Основной упор в работе сделан на методологии исследования: сопоставление результатов, полученных с использованием весьма различных (по используемым метрикам и принципам разбиения) методов анализа данных: порядково-инвариантной паттерн-кластеризации (с корректировкой на центроиды) и иерархической кластеризации.

## **4 Исследования моделей по оптимальному размещению объектов и привлечению частных инвестиций**

Эта часть отчета посвящена результатам исследования моделей по оптимальному размещению объектов и привлечению частных инвестиций. Указанное исследование осуществлялось в следующих направлениях:

- теоретико-игровой подход к оценке экономической целесообразности создания нового грузового транспортного узла;
- исследование задачи оптимизации размещения в городских районах станций по зарядке и по замене батарей для питания электромобилей.

### **4.1 Теоретико-игровой подход к оценке экономической целесообразности создания нового грузового транспортного узла**

#### **4.1.1 Постановка основных задач исследования**

Любое обсуждение экономического развития региона обычно включает в себя обмен идеями о совершенствовании системы грузоперевозок. Это особенно актуально, если это улучшение частично мотивировано определенными преимуществами географического положения региона. Любые региональные грузы нуждаются в транспортных услугах — между пунктами в пределах региона, между пунктами в других регионах и пунктами в регионе, а также при их транзите через территорию региона. Именно такие преимущества в конечном итоге позволяют обеспечить более быструю и дешевую доставку большинства грузов как результат использования транспортных услуг в регионе. Поэтому совершенствование транспортной инфраструктуры региона является одной из важнейших задач любой региональной администрации, особенно тех регионов, где это совершенствование позволяет привлечь как частные, так и государственные инвестиции и создать новые рабочие места.

Рассмотрим для примера некоторые пункты назначения грузов для набора грузов расположенных в той части региона, которая находится недалеко, скажем, от речного порта  $A$ , и пусть все эти грузы в настоящее время доставляются из точек их отправления в пункт назначения грузовиками. Да-

лее, пусть пункт отправления каждого из этих грузов будет близко к конкретному речному порту (или к ряду конкретных речных портов) и может быть доставлен в этот конкретный речной порт (или в некоторые из этих речных портов) грузовиками. Наконец, предположим, что каждый груз мог бы быть доставлен из порта *A* в пункт назначения грузовиками, если бы порт *A* был пунктом отправления для этого груза. Тогда можно рассмотреть следующую схему перемещения каждого груза (или части грузов) между его (их) пунктом (пунктами) отправления и его (их) пунктом (пунктами) назначения: Конкретный груз (конкретные грузы) перемещается (перемещаются) (а) из его (их) пункта зарождения в ближайший речной порт *B*, (б) перемещается на речных судах из этого ближайшего порта *B* в речной порт *A*, и (с) перемещается из порта *A* в пункт (пункты) назначения груза (грузов).

Все разумные рассуждения, естественно предполагают, что перевозки груза (грузов) на речном транспорте или уже осуществляются или могут быть организованы между каждым из вышеупомянутых ближайших речных портов и портом *A*.

Для того чтобы региональная администрация рассматривала подобные схемы даже для некоторой группы грузов, территориально "тяготеющей" к какому-либо одному речному порту, ей необходимо иметь в своем распоряжении инструмент, позволяющий ей оценить экономическую целесообразность этих схем, а также экологические и социальные последствия, которые могут быть связаны с их практической реализацией. При этом должна обеспечиваться возможность рассмотрения модернизации действующего речного порта, строительства нового речного порта *C* в регионе, участия в строительстве нового речного порта *D* вне региона и т.д. исходя из (а) объемов потоков грузов, которые предполагается перемещать между пунктами зарождения и назначения грузов, входящих в эту группу грузов, (б) ожидаемых тарифов на транспортировку исходя из прогнозируемых рыночных значений этих тарифов как на перевозки грузов на речном, так и на автомобильном транспорте, а также (с) ожидаемых затрат на строительство и эксплуатацию транспортного

узла, который будет организован в месте расположения речного порта А, или в месте расположения нового речного порта С (или в обоих географических пунктах).

Ожидаемые значения всех этих параметров позволили бы администрации региона количественно оценить доход, который проекты перевалки, реализующие каждую из вышеупомянутых схем, могут генерировать при всех возможных сценариях и решить, может ли разработка какого-либо из этих проектов быть интересной для региона.

Ясно, что значения объемов потоков грузов через потенциальные речные транспортные узлы, тарифов на транспортные услуги по перемещению грузов и на обслуживание этих транспортных узлов и подъездных путей к ним, а также затрат на строительные работы, связанные с созданием и модернизацией транспортных узлов и подъездных путей к ним, имеют вероятностный характер. В то же время, исходя именно из их оценок и вырабатываются возможные стратегические решения, включая финансовые, которые принимает региональная администрация при анализе вариантов развития региона, включая социальное развитие.

*Объемы грузопотоков.* Представляется практически невозможным надежно предсказать значения этих чисел, например, на ближайшее десятилетие или два. Однако оценки нижних границ этих объемов могут быть получены из анализа динамики а) экономического развития, как региона, так и его соседних территорий, включая открытие на них новых и разработку уже обнаруженных полезных ископаемых, б) внутренней и международной торговли в рамках трех вышеупомянутых схем для перемещения грузов по территории региона (ввоз грузов в пункты региона, вывоз грузов из региона, и транзит грузов через регион) и в) миграционных процессов, рождаемости и смертности. Таким образом, если, например, номенклатура грузов, перемещаемых через конкретную географическую точку, считается состоящей из конечного числа наименований, а ожидаемые объемы этих грузов образуют вектор в конечно-



мерном пространстве, размерность которого совпадает с числом наименований в номенклатуре, то нижние оценки этих объемов за определенный период времени можно описать векторным неравенством, в котором вектор этих объемов ограничен снизу некоторым вектором, компонентами которого являются нижние границы значений объемов соответствующих грузопотоков.

*Тарифы на услуги, связанные с перевозкой грузов в регионе и через регион.* Эти тарифы во многом зависят от а) развития видов грузового транспорта функционирующего в этом регионе, б) количества игроков на рынке транспортных услуг (компаний, предлагающих транспортные услуги на рынке этих услуг в регионе); в) объемы грузопотоков, предлагаемых к перевозке отправителями грузов. В общем случае ожидаемые тарифы на грузоперевозки на региональном рынке можно описать системой линейных ограничений балансного типа [180]- [184]. Эту систему можно записать в векторно-матричной форме в виде матричного неравенства вида  $Bb \leq h$ , где  $B$  – матрица с действительными элементами, а  $b$ ,  $h$  – векторы соответствующих размеров с компонентами, являющимися действительными числами, причем  $b$  является вектором тарифов на услуги по перевозке грузов.

*Стоимость строительных услуг.* Стоимость строительных услуг, связанных с созданием нового транспортного узла (или новых транспортных узлов) и подъездных путей к нему (к ним) существенно зависит от размеров строящегося комплекса (строящихся комплексов). В случае же развития уже функционирующего транспортного узла, объем этих работ зависит от а) объемов грузопотоков которые будут перемещаться через этот узел, б) типов грузов, которые будут перемещаться через этот узел, в) санитарных и технологических требований, подлежащих выполнению и связанных с нахождением некоторых типов грузов на территории транспортного узла, и г) естественных ограничений на площадь территории, которую может занимать этот транспортный узел. Существуют подходы к определению оптимальной структуры транспортного узла, оптимальных размеров отдельных его подразделений и

оптимальной последовательности всех строительных работ, которые будут там выполняться [185] - [188].

Ясно, что если бы, например, число контейнеров всех размеров, прохождение которых ожидается через контейнерный терминал в течение какого-либо промежутка времени (например, года), было известно, то размер контейнерного терминала можно было бы приблизительно оценить как некоторое число (числа), пропорциональное этому (этим) размеру (размерам) с определенным коэффициентом пропорциональности (определенными коэффициентами пропорциональности). В свою очередь, размер конкретного контейнерного терминала определил бы перечень строительных работ, которые необходимо выполнить для строительства этого терминала, включая объем каждой работы из этого перечня. Поскольку, как правило, стоимость каждой работы из их перечня может быть известна только в определенных диапазонах, естественно предположить, что стоимость строительства терминала может быть описана билинейной функцией двух векторных аргументов [189]. Компонентами первого векторного аргумента будут ожидаемые объемы всех работ из перечня, а компонентами второго векторного аргумента будут ожидаемые затраты на эти работы.

Аналогичные рассуждения применимы ко всем типам грузов, которые, как ожидается, будут перемещаться через транспортный узел. Таким образом, можно сделать вывод, что стоимость всех строительных услуг в транспортном узле может быть описана билинейной функцией в виде скалярного произведения двух векторных аргументов. Первый векторный аргумент образован вектором, компонентами которого являются объемы строительных работ, находящихся в указанном выше перечне таких работ, подлежащих выполнению для каждого объекта, входящего в состав транспортного узла. Эти объемы строительных работ зависят от объемов грузов, которые предполагается обрабатывать в этом узле. Второй векторный аргумент образован вектором, составляющими которого являются затраты на все эти строительные работы, аналогично тому, что было сказано выше для контейнерного терминала.

Здесь набор допустимых значений для первого векторного аргумента может быть описан системой линейных уравнений. Эта система связывает векторы объемов всех строительных работ в транспортном узле и векторы объемов всех транспортных потоков, которые предполагается перегрузить через транспортный узел в течение конкретного периода времени. Система может иметь вид  $E(p)w - Hx = 0$ , где  $w$  и  $x$  векторы в  $p$ -мерном и в  $n$ -мерном пространствах, соответственно, причем:

а)  $p$  – число позиций в номенклатуре всех выполняемых строительных работ;

б)  $E(p)$  – единичная матрица размера  $p$ ;

в)  $H$  – матрица размера  $p \times n$  с неотрицательными элементами;

г)  $x$  – вектор с компонентами, равными объемам грузов, которые предполагается перегрузить через транспортный узел, для которых выполняется множество линейных неравенств, в том числе множество неравенств  $E''(x - y) \leq 0$ , где  $E''$  – единичная матрица размера  $n$ . (Как и в предыдущем случае,  $y$  –  $n$ -мерный неотрицательный вектор ожидаемых объемов грузов в регионе для перемещения в течение определенного периода времени, а  $n$  – число наименований в номенклатуре предполагаемых к перемещению грузов.)

Набор допустимых значений для второго векторного аргумента с билинейной функции может быть описан системой линейных неравенств вида  $\Psi c \leq t$ , где  $c$  –  $p$ -мерный неотрицательный вектор,  $\Psi$  – матрица размера  $r \times p$  с вещественными элементами и  $t$  –  $r$ -мерный вектор. Здесь компонентами этого вектора  $c$  являются стоимости строительных работ, которые предполагается выполнить в рамках строительства транспортного узла.

Целью проведенного исследования было показать, что для оценки экономической (финансовой) целесообразности разработки грузового транспортного узла в некоторой географически привлекательной точке региона можно сформулировать и решить оптимизационную задачу в робастной постановке. Именно, целью было показать, что исходя из приведенных выше рассуждений об ожидаемых а) объемах грузопотоков, б) тарифах на транспортные услуги,

и в) затратах на строительные услуги, можно определить есть ли у такого узла (при определенной его мощности) какая-либо технически осуществимая возможность принести приемлемый доход в бюджет региона в течение разумного периода времени. Например, годовой доход, который можно ожидать от функционирования транспортного узла, может считаться приемлемым, если он (по крайней мере) покрывает инвестиции, необходимые для развития этого узла в такой срок, который представляется разумным тем, кто рассматривает возможность инвестировать капитал в этот проект. При рассмотрении этой возможности следует иметь в виду, что а) транспортный узел должен конкурировать с другими видами транспортных услуг, доступных на рынке транспортных услуг по любым ожидаемым рыночным транспортным тарифам, б) цены на все строительные работы также определяются рынком строительных услуг и в) вектор объемов всех необходимых строительных работ  $w$  и вектор объемов грузов  $x$ , которые предполагается перегружать через транспортный узел, являются решениями системы векторных линейных уравнения вида  $E(p)w - Hx=0$ , где  $p$  – натуральное число, а  $E(p)$ ,  $H$  – матрицы упомянутые выше.

Как показано в настоящем исследовании, эта робастная оптимизационная задача может быть сформулирована как задача отыскания максимума максимина функции, зависящей от четырех векторных аргументов. Каждый из этих векторных аргументов принадлежит некоторому своему выпуклому многограннику, значения двух из этих четырех векторов связаны системой линейных уравнений, а сама функция (максимум максимина которой подлежит отысканию) представляет собой разность двух квадратичных функций. В проведенном исследовании доказано, что решение этой задачи сводится к решению некоторой вспомогательной задачи квадратичного программирования, одна из постановок которой предполагает отыскание наиболее невыгодных тарифов на рынке транспортных услуг. Последняя задача формулируется как задача билинейного программирования, решение которой сводится к решению пары двойственных задач линейного программирования, как это показано в

[189]. В математической формулировке упомянутой выше задачи робастной оптимизации система ограничений для векторного аргумента, описывающего рыночные тарифы, содержит специальную систему линейных неравенств, совместность которой является достаточным условием конкурентоспособности перевалочного тарифа в транспортном узле. Совместимость этой системы неравенств легко проверить методом, предложенным в [190].

В исследовании сформулировано и доказано достаточное условие конкурентоспособности тарифов на услуги, оказываемые в транспортном узле, а также математически сформулированы все рассматриваемые в нем прикладные задачи и доказаны соответствующие утверждения. Как уже указывалось выше, продемонстрирована возможность формулировки задачи об оценке финансовой целесообразности строительства и эксплуатации нового транспортного узла в виде задачи нахождения максимума максимина функции двух векторных аргументов, являющейся разностью двух билинейных функций и показана возможность отыскания ее решения из решения вспомогательной задачи квадратичного программирования. Предложен обзор результатов работ, опубликованных разными авторами, по задачам, близким по постановке к задачам, рассмотренным в настоящем исследовании, показана связь между настоящим исследованием и работами, опубликованными авторами ранее [191], [192], а также обсуждаются направления дальнейших исследований в рассмотренной области, которые представляются авторам целесообразными именно для специалистов по прикладной математике (хотя прикладные аспекты рассмотренных проблем традиционно изучаются лишь специалистами в области управления транспортом).

Проведенное исследование выполнялось совместно с Г. Фединым, бывшего сотрудника МЦАВР, который успешно защитил кандидатскую диссертацию в октябре 2020 года, и профессором Принстонского университета А. Корнхаузером, с которым совместные работы в области управления транспортом ведутся более 30 лет.

#### 4.1.2 Краткий обзор состояния работ в области исследования

Проблемы конкурентного местоположения объектов вообще и транспортных систем в частности широко изучаются уже в течение нескольких десятилетий. Впервые такая проблема была сформулирована в [193] как проблема отыскания размещения двух рыночных игроков с равномерным распределением потребителей их услуг, в которой автор предположил, что оба участника рынка принимают решения об их местонахождении и предлагают цены на свои услуги одновременно. С момента публикации этой статьи была проделана большая работа на указанную тему и опубликовано много статей. В частности, [194] одним из первых сформулировал задачу о конкурентном размещении конечного числа объектов для взвешенных графов произвольного типа, где он рассмотрел рынок с одним лидером и одним последователем, на котором оба пытаются максимизировать сумму весов узлов, для которых их объект является ближайшим. Так были сформулированы две *NP*-трудные задачи: В первой из них лидер решает, где разместить  $p$  объектов, предполагая, что затем последователь решит, где разместить  $r$  объектов, в то время как во второй задаче ведомый решает, где разместить  $r$  объектов, зная, что лидер уже разместил  $p$  объектов. Обзор статических задач конкурентного месторасположения объектов был предложен в [195].

Задача последовательного размещения нескольких объектов была далее сформулирована в [196] в виде игры двух лиц, где авторы рассмотрели последовательную пошаговую модель процесса. На каждом шаге этого процесса один игрок принимает решение только об одном месте расположения объекта, а игра продолжается до тех пор, пока оба игрока не определяют расположение всех интересующих их объектов. Последняя публикация, касающаяся задач этого типа [197], сосредоточена на последовательных проблемах конкурентного размещения, в которых решение о размещении объектов сначала принимает лидер рынка, а затем его последователь [198] сделали еще один шаг вперед в исследовании задач последовательного размещения объектов, предположив, что число новых предприятий последователя заранее не определено и

предложив робастный подход к определению местоположения объектов в задаче оптимизации, решаемой лидером рынка.

Математические модели, с помощью которых различные аспекты проектирования объекта и проблемы его размещения изучаются совместно предложены в [199], [203]. В частности, в [199], участник рынка стремится определить местоположение и привлекательность каждого объекта, который будет открыт в регионе, принимая во внимание расположение существующих объектов конкурента. Конкурент может отреагировать на действия участника путем корректировки привлекательности своих объектов. Цель каждого игрока состоит в максимизации собственной прибыли. в зависимости от доли игрока на рынке. Эта доля рынка рассчитывается с помощью правила гравитации и пропорциональна привлекательности объекта и обратно пропорциональна расстоянию между объектом и пользователем. Авторы сформулировали рассматриваемую задачу как двухуровневую нелинейную задачу смешанного математического программирования. Используя условия оптимальности Каруша-Куна-Таккера, авторы преобразовали эту двухуровневую задачу нелинейного программирования в эквивалентную одноуровневую нелинейную задачу смешанного математического программирования. При решении примеров рассмотренной ими проблемы с числом вершин равным 30 и с числом узлов до 5, авторы использовали GMIN-aBV алгоритм, реализующий метод глобальной оптимизации, предложенный в [204].

В [203] автор рассматривает задачу, решаемую последователем при выборе им оптимальных расположений заранее определенного числа магазинов и определения подходящей цены на продаваемый в них однородный продукт с целью максимизации прибыли от продажи этого продукта в конкурентной среде. В этой задаче предполагается, что последователь принимает во внимание расположение магазинов конкурента (лидера рынка), продающих тот же товар, и цены в них. Однако, в отличие от подхода, предложенного в [199], действующий на рынке конкурент не реагирует на активность последователя изменением собственной цены. Другое (хотя и аналогичное предложенному в

[199] предположение, в котором считается, что функция полезности потребителя зависит от цены (привлекательность) и стоимости поездки и описывает поток клиентов, направляющихся в магазины, используя полиномиальную логит-модель. Для решения полученной таким образом задачи нелинейного смешанного программирования, автор делит проблему на две части. В одной из них определяется местоположение магазинов, а в другой цена на продаваемый продукт, причем каждая из двух частей задачи решается отдельно с использованием нескольких эвристик. Для решения первой части задачи, относящейся к отысканию местоположения магазинов, автор применяет greedy алгоритм, процедуру запрета поиска и генетический алгоритм. Для решения второй части задачи, относящейся к отысканию цены, автор использует метод градиентного спуска, метод градиентного спуска с несколькими случайными начальными точками и алгоритм следования по пути. Предложенные эвристики были протестированы на наборе сгенерированных проблем с 40 новыми магазинами.

Другие исследователи рассматривают игры ценообразования как часть задачи местоположения. Например, [205] изучают игру двух лиц, в которой игроки выбирают свои цены и свое местоположение в то же самое время. Авторы доказывают, что в рассматриваемой ими последовательной игре, последователь всегда зарабатывает больше чем лидер. Похожие игры ценообразования изучаются в работах [206], [208].

Хотя конкуренция в выборе местоположения была изучена подробно, конкуренция между действующим на рынке оператором и потенциальным участником узловых сетей с парами происхождения-назначение (OD) изучены гораздо меньше. По-видимому, работу [209] следует считать первой статьей, посвященной этому типу проблемы конкурентного размещения, в которой авторы предложили математическую постановку задачи ведомого (последователя), где некоторая транспортная компания уже работает в распределенной узловой сети, а новая компания хочет выйти на тот же рынок. В предложенной там математической модели, последователь захватывает 0,5, 0,75 или даже



весь поток между парами OD, в зависимости от разницы между тарифами последователя и действующего оператора, которые установлены для потребителей. Для решения задачи, сформулированной на основе предложенной авторами математической модели с использованием набора данных AP, с 20-, 25-, 40- и 50-узловыми сетями, ими была разработана эвристика. Однако, несмотря на то, что сама по себе эвристика корректна, ее реализация была оспорена в [210]. Задача, аналогичная [209], рассмотрена в [211], где авторы изучают проблему последователя (ведомого) применительно к проблеме расположения аэропорта. Основное отличие предложенной ими модели от модели, предложенной в [209], состоит в способе моделирования предпочтения клиентов. Именно, в [211] непрерывные предпочтения клиентов моделируются с помощью функций полезности, подобных гравитации. Предпочтения клиента зависят от основной привлекательности объекта (аэропорта), тарифов и времени в пути. Полученная задача нелинейной целочисленной оптимизации была решена с помощью эвристики, предложенной авторами, и модель и эвристика были протестированы на наборе данных AP с 25 узлами и числом новых концентраторов, доходящим до шести.

В отличие от авторов работы [209], автор работы [212] рассматривает задачу, в которой две фирмы решают, где разместить узлы и какие пары OD должны использоваться. Автор предполагает, что лидер принимает решение, первым решая двухуровневую задачу, в которой проблема второго уровня является проблемой последователя (ведомого). Достижимая доля рынка определяется на основе логит-функции с заранее заданной полезностью, которую можно интерпретировать как тарифы. В предложенную автором математическую модель, лежащую в основе математической постановки рассматриваемой им задачи, автор включает ограничения на пороговые значения потока, которые используются для определения обслуживаемых пар OD. Проблема была решена как методом полного перебора, так и с помощью предложенной автором greedy (жадной) эвристики. Подобно [212], авторы [213] изучают проблему, в которой две фирмы последовательно решают, где разместить узлы, а

затем клиенты выбирают одну из фирм в зависимости от уровня предлагаемых ими услуг или тарифов. Фирма, предлагающая лучший уровень обслуживания, захватывает весь поток между определенной парой вершин сети. Авторы предлагают две математические модели и формулируют на их основе две задачи. Для лидера формулируется двухуровневая задача, а задача о максимальном покрытии формулируется для последователя, исходя из предположения, что узлы лидера уже были размещены. В обеих задачах целью каждой фирмы является максимизация своей доли рынка. Доказано, что обе задачи являются *NP*-полными, и авторами предложены эвристики для их решения. Предложенные модели, и алгоритмы были протестированы и проанализированы на наборах данных CAB и TR.

В [214] авторы рассматривают конкурентоспособное расположение узла с ценовой конкуренцией. Эта проблема описывает ситуацию, в которой новая компания а) строит свою собственную сеть узлов и лучей (hub-and-spoke систему) с возможными дополнительными прямыми соединениями между клиентами, и б) устанавливает тарифы. Логит-функция используется для моделирования предпочтений клиентов как в отношении уже действующей на рынке компании, так и в отношении последователя (в зависимости от его ценовой стратегии). Рассматриваются две целевые функции последователя – прибыль и рыночная доля. Эти целевые функции максимизируются за счет: а) выбора наилучшего расположения узлов и топологии сети и б) применения оптимальной цены. В статье авторы формулируют нелинейную задачу смешанного математического программирования, которую они решают с помощью генетического алгоритма, и предлагают результаты тестов и вычислительных экспериментов, использующих наборы данных CAB с 25 узлами.

## **4.2 Исследование задачи оптимизации размещения в городских районах станций по зарядке и по замене батарей для питания электромобилей**

### **4.2.1 Постановка основных задач исследования**

Быстрое развитие новых технологий для получения и доставки электроэнергии, а также экологические преимущества использования электроэнергии, порождают как научные, так и практические проблемы. В частности, подлежат изучению проблемы, связанные с поиском оптимальных структур систем снабжения электричеством объектов и технологических комплексов, обслуживающих различные сферы жизни и бизнеса. Актуальность разработки систем размещения сети зарядных станций, снабжающих электричеством движущиеся в городе электромобили, объясняется: а) проникновением электромобилей в повседневную жизнь миллионов людей во всем мире; б) очевидными экологическими преимуществами этих автомобилей по сравнению с автомобилями, работающими на ископаемом топливе.

Цель настоящего исследования – предложить подход к решению части этой проблемы. Именно, предполагается, что требуется оценить стоимость создания указанной выше сети исходя из:

- потребности в электроэнергии в сети, во многом зависящей от количества электрических транспортных средств, которые будут использоваться в городе в течение любого набора месяцев (или лет) в будущем;
- существующих требований к структурированию этой сети, продиктованных как региональными, так и федеральными постановлениями (например, чтобы зарядная станция в сети находилась бы в каждом, скажем, участке города в пределах круга определенного радиуса);
- ожидаемого производства мощностей всех видов возобновляемых источников энергии, которую можно использовать на зарядных станциях сети (в зависимости от погодных условий в каждой части города);

- стоимости оборудования (электростанций), которое будет установлено на каждой зарядной станции сети для обеспечения водителей электротранспорта (а в будущем и автоматизированных электромобилей) электричеством;
- стоимости технического обслуживания электростанций в сети;
- стоимости эксплуатации каждой зарядной станции сети.

Проблема оптимизации размещения зарядных станций для электромобилей в крупном городе служит примером из быстро растущей группы проблем, связанных с финансированием крупных проектов на развитие систем распределения электроэнергии в целом.

На первый взгляд может показаться, что указанная проблема принадлежит к области транспорта больше чем к области электроснабжения. Однако ясно, что как только либо спрос на электричество на каждой зарядной станции из сети определен, или хотя бы диапазон этого спроса был оценен, это уже более не так. Именно, часть проблемы развития вышеупомянутых сетей зарядных станций становится полностью идентичной разработке систем электроснабжения набора жилых кварталов или набора коммерческих предприятий. Эта часть системы городского электроснабжения еще достаточно сложна и многие владельцы электромобилей предпочитают заряжать свои электромобили дома в ночное время, чтобы сэкономить деньги на зарядных станциях из сетей. Таким образом, следует разумным образом организовать снабжение электричеством тех представителей городского населения, которые хотели бы иметь электричество для зарядки своих электрических транспортных средств в местах их проживания или очень близко к ним расположенных, в дополнение поставкам электричества, покрывающим их повседневные бытовые нужды. Последняя проблема вряд ли может рассматриваться как представляющая интерес только для транспортных систем.

Технологически, каждая конкретная городская станция для зарядки электромобиля электричеством может использовать различное оборудование. Тип, размер и число единиц оборудования любого типа и размера зависят от

размера и структуры потока электромобилей, обслуживаемых на этой зарядной станции, который и определяет потребность в электроэнергии на ней. При этом, однако, в определенной степени, этот поток зависит и от стоимости зарядки электромобиля, что, в свою очередь, зависит от стоимости подачи на станцию электроэнергии из соответствующей региональной сети электроснабжения (или из нескольких региональных сетей электроснабжения) и из других источников электроэнергии (например, со станций, на которых осуществляется выработка электроэнергии из возобновляемых источников энергии).

Предположим, что городская администрация и/или компания, называемая далее заинтересованным юридическим лицом, рассматривает развитие и функционирование сети зарядных станций для электромобилей в городе в намерении сделать обеспечение жителей города-владельцев электромобилей, а также иных находящихся в городе лиц, прибыльным бизнесом. Далее предположим, что это заинтересованное юридическое лицо исходит из экспертных заключений и оценки а) ожидаемого спроса на электроэнергию от электромобилей на каждой зарядной станции города, и б) ожидаемых цен на электроэнергию в любой региональной электросети и в любых других источниках энергоснабжения электричеством зарядных станций города. Кроме того, предположим, что вышеупомянутая информация будет известна каждый час в каждый 24-часовой период времени, скажем, в каждый месяц каждого из четырех сезонов года. Наконец, предположим, что это заинтересованное юридическое лицо намеревается получить количественную оценку размера инвестиций, необходимых для развития сети зарядных станций в городе и их прибыльной эксплуатации, в соответствии с указанными выше и некоторыми другими ограничивающими, хотя и проверяемые предположениями.

Проблема оценки размера этих инвестиций и составляет предмет исследования в настоящей работе.

Представленный обзор подходов к математическому моделированию проблем, близких к сформулированной, и краткое описание использования

этих подходов при формализации соответствующих прикладных задач позволяют заключить, что математическая модель, предложенная в [215] для задачи оптимального размещения зарядных станций для электромобилей на участке скоростной автомобильной магистрали, может быть успешно использована и для задачи размещения таких станций в крупных городах. Именно, предложенная в [215] математическая модель помогает формализовать рассматриваемую проблему как задачу отыскания максимина функции, являющейся суммой нескольких линейных и двух билинейных функций, большого размера с линейными ограничениями, смешанными переменными и некоторыми дополнительными (по сравнению с задачей, рассмотренной в [192]) ограничениями. Как и задачу, рассмотренную в [192], оказывается возможным рассматривать эту задачу как задачу нелинейного программирования со смешанными переменными, и решение этой задачи нелинейного программирования большого размера удастся свести к решению задачи смешанного математического программирования с целевой функцией и системой ограничений, имеющими линейную структуру, что доказывается совершенно аналогично тому, как это сделано в [215].

Возможность быстро и точно решать крупномасштабную задачу, рассматриваемую в настоящем исследовании, при любых исходных данных и экспертных оценках, с использованием широко доступных стандартных пакетов программ для решения задач математического программирования, чрезвычайно важна. Именно эта возможность превращает сочетание предложенного формализованного описания проблемы на основе использования упоминавшейся выше математической модели и эти стандартные пакеты программ в мощный аналитический экономический инструмент. Такой инструмент может помочь региональным администрациям, местным властям и другим заинтересованным юридическим лицам вести переговоры с частным бизнесом о возможных частных инвестициях в развитие сети зарядных станций для электромобилей в городах, например, в рамках государственно-частного партнерства любой формы.

Этот инструмент также полезен и для потенциальных частных инвесторов, которые рассматривают возможные инвестиции в развитие сети в переговорах с заинтересованным юридическим лицом, а также для местных органов власти при анализе общих финансовых вопросов, связанных с развитием и управлением городской сетью зарядных станций для электромобилей. Также как и при разработке сети зарядных станций для электромобилей, движущихся вдоль скоростных магистралей, следует учитывать возможность питания зарядных станций электричеством, получаемым из возобновляемых источников энергии, включая солнечные батареи и ветряные турбины.

Также как и в задаче размещения зарядных станций для электромобилей вдоль скоростных магистралей, следует учитывать разницу в ценах на электроэнергию, которые электросети предлагают своим клиентам в течение 24 часов. Как уже упоминалось ранее, известно, что эти цены в ночные часы могут быть существенно ниже цен на электроэнергию, предлагаемых сетями электроснабжения в дневное время, особенно в часы пик. Таким образом, следует рассмотреть возможность использования а) больших систем хранения электроэнергии, покупаемой у электрических сетей в ночное время, которые могут обеспечить зарядные станции в системе зарядных станций города, и б) систем хранения электроэнергии малых и средних размеров который может располагаться, например, на некоторых зарядных станциях города [216]. Математические модели, подобные предложенным в [215] и в настоящем исследовании, могут быть использованы при анализе вариантов развития городских сетей зарядных станций для электромобилей, учитывающих возможности развертывания: а) систем хранения электроэнергии как небольших (устанавливаемых на зарядных станциях таких городских сетей), так и больших (общих систем хранения электроэнергии); б) возобновляемых источников энергии, таких как солнечные батареи и ветряные турбины, и в) обоих источников энергии. Хотя такие модели не являлись предметом настоящего исследования, при их разработке могут оказаться полезными результаты, полученные автором в [216], [217].

#### 4.2.2 Краткий обзор подходов к математическому моделированию проблем, близких к проблеме, рассматриваемой в настоящем исследовании

Математическая модель для формализации проблемы размещения зарядных станций для электромобилей в большом городе в виде двухкритериальной задачи предложена в [218]. Авторы рассматривают а) способы инвестирования в соответствующие крупномасштабные проекты, б) роль государства в реализации проектов, и в) соотношения между мощностями зарядных станций и их расположением от сетей электроснабжения. Они предлагают свои соображения по финансовым и организационным вопросам, основанные на проведенных численных расчетах с использованием предложенной ими математической модели применительно к району города Пекин. Модель, являющаяся моделью с линейными ограничениями и смешанными переменными, помогает авторам изучить зависимость первоначальных (авансовых) инвестиций на эксплуатационные расходы и покрытие услуг, которое система зарядных станций может предоставить.

Два метода для определения расположения станций замены батарей для электромобилей предлагаются и сравниваются в [219]. Методы, которые называются их авторами аналитическими и метаэвристикой Искусственных пчелиных колоний, были протестированы на системе тестирования автобусов IEEE, и результаты тестирования показали близость результатов их использования друг к другу. Кроме того, авторы сообщают об использовании метаэвристики Искусственных пчелиных колоний для определения размера станций для замены батарей для электромобилей в качестве метода определения объемов производства распределенных генераторов электроэнергии для электромобилей. По мнению авторов, определение оптимального размера и оптимальной производительности одновременно обеспечивает минимальные потери мощности батарей в размещенных станциях замены батарей и высочайшую стабильность их работы.



Проблема определения, как размера, так и расположения станций быстрой подзарядки электроэнергией (аккумуляторных) электрических транспортных средств на сети автомобильных дорог с недостаточно развитой существующей инфраструктурой таких станций рассматривается в [220]. В своих рассмотрениях авторы исходят из ограниченного бюджета на развитие проекта и конкретных возможностей зарядных станций для обслуживания электромобилей. Принимая во внимание возможные стратегии водителей таких транспортных средств по их зарядке электроэнергией, авторы рассматривают равновесное распределение транспортных потоков при недостаточной существующей инфраструктуре. Они вычисляют это равновесное распределение на основе итеративного подхода, предложенного в статье.

Рассматриваемая ими задача формулируется как задача смешанного математического программирования, в которой а) конкретные задания водителям выдаются с учетом расположения зарядных станций и б) общий поток автомобилей,двигающихся по сети дорог, является максимальным. Три алгоритма для решения этой задачи смешанного математического программирования предлагаются, и проблема рассматривается для конкретной сети автомобильных дорог в Китае.

Проблема оптимального распределения станций для быстрой подзарядки аккумуляторных электромобилей как шаг в направлении развития инфраструктуры системы таких станций подзарядки в стране рассматривается в [221] на основе данных США о дальних поездках. Оптимальная структура понимается авторами как структура, которая максимизирует число таких дальних поездок, которые выполняется водителями, эксплуатирующими аккумуляторные электрические транспортные средства, и задача отыскания такой структуры формулируется авторами как задача смешанного математического программирования. Модификация известной математической модели проточного размещения зарядных станций рассматривается и формулируемая на основе этой модели задача отыскания расположения таких станций решается с помощью метода ветвей и границ. Результаты расчета для конкретных данных,

когда определенное число зарядных станций (от 50 до 250) определяется для определенного числа миль (от 60 до 250), проходимых электромобилем за завершённую поездку, представлены авторами в их статье. Авторы предлагают рекомендации по расположению станций быстрой зарядки для аккумуляторных электромобилей.

Четырёхэтапный подход к решению задачи оптимального размещения станций подзарядки электромобилей в большом городе предложен на базе (CIS) географической информационной системы в [222]. В рамках этого подхода а) данные из геоинформационной системы используются для присвоения оценки участку, где станция по зарядке электромобилей потенциально может быть расположена, б) выбранные критерии (число которых равно 15) приоритизируются так называемым нечетким аналитическим иерархическим процессом, и в) потенциальные места для расположения зарядных станций ранжируются с использованием специальных методов. Подход был протестирован на данных, относящихся к городу Анкара, и, согласно авторам статьи, использование их результатов позволило предложить лучшие места для расположения 12 зарядных станций (с точки зрения выбранных критериев), чем те, которые используются в настоящее время.

В [223] методы имитационного моделирования предложено использовать для анализа трех параметров, связанных с электрическими транспортными средствами. Именно, исследуются а) потенциал роста продаж аккумуляторных электромобилей в США, б) пробег, который может покрыть электромобилями, и в) уровень выбросов парниковых газов электромобилями. Эти параметры рассматриваются с точки зрения влияния, которое развитие общественной инфраструктуры зарядки электромобилей в США может иметь на значения этих параметров. Авторы сообщают о двух подходах к реализации имитационного моделирования (сценарный анализ и параметрический анализ), и они приходят к некоторым выводам о наиболее эффективных быстрых зарядных устройствах в местах крупного развертывания общественных зарядных устройств. Именно, согласно авторам, такие зарядные устройства а)

увеличивают продажи аккумуляторных электромобилей, б) увеличивают пробег электромобилей, и в) снижают выбросы парниковых газов.

Авторы [224] рассматривают проблему снижения мощности в сети электроснабжения, обеспечивающей электроэнергией некоторое множество зарядных станций, вызванного одновременным подключением большого числа электромобилей в эту сеть для подзарядки. Они предлагают метод, в рамках которого оптимальное распределение активной и реактивной мощности между несколькими зарядными станциями определяет стратегию сотрудничества этих станций. Авторы предлагают результаты расчетов, показывающие уменьшение потерь мощности за счет использования их метода, которые они демонстрируют на примере распределительных автобусных тест-линий. Также, они оценивают ожидаемый годовой доход, который могут иметь инвесторы, вкладывая в строительство и эксплуатацию дополнительных зарядных станций в соответствии с рекомендациями их метода.

Расчеты числа электромобилей, приходящихся на одну общественную станцию быстрой зарядки электромобилей с высокой мощностью, ожидаемую быть востребованной в будущем, представлены в [225]. Эти расчеты проводятся для оценки потребностей в инфраструктуре таких зарядных станций для транспортных средств. Авторы используют набор данных из Швеции и Норвегии и математическую модель типа модели очередей для оценки этого отношения. По их расчетам с использованными данными они предполагают, что ожидаемое отношение будет близко к уровню существующему в системах обычной заправки. Однако, их оценки зависят не только от использованных данных, но и от а) размера батарей в электромобиле, и б) уровня мощности на зарядных станциях, который, как отмечают авторы, может увеличиваться в будущем.

Задача размещения набора зарядных станций, которые могут обслуживать работающие на подключении гибридные электромобили, сформулирована и рассмотрена в [226]. Эта формулировка является обобщением форму-

лировки задачи определения местоположения перезаправки потока, где максимальный поток автомобилей, который может обслуживаться конечным числом зарядных станций, находится под определенными ограничениями, накладываемые на типы обслуживаемых электромобилей. Целевой функцией поставленной задачи является число миль, которые проезжает гибридный автомобиль на электроэнергии. Эта функция максимизируется при существующих соотношениях между стоимостью электроэнергии и стоимостью бензина, и эта максимизация эквивалентна минимизации общих транспортных расходов, связанных с транспортным средством. Для решения рассматриваемой проблемы предлагается модификация алгоритма декомпозиции Бендерса. По мнению авторов, результаты расчетов, о которых они сообщают, свидетельствуют о том, что, предложенный ими алгоритм эффективно решает проблему при реальных ее размерах. Они также сообщают, что этот алгоритм эффективно решает задачу определения местоположения перезаправки потока, которую рассмотренная ими задача обобщает.

Сравнение подхода многоцелевой оптимизации к размещению общественных станций для быстрой зарядки электромобилей на основе единого оптимизационного и подхода на основе численных расчетов предлагается в [227]. Именно, результаты работы многоцелевого эволюционного алгоритма (усиления эволюционного алгоритма Парето-II) для решения задач многокритериальной оптимизации сравниваются с результатами работы одноцелевого алгоритма (улучшенного генетического алгоритма).

Многопериодная модель местоположения многолучевой дозаправки, отражающая динамику изменений в сети общественных зарядных станций, вызванная ростом рынка электромобилей, предлагается в [228]. На основе этой модели авторы формулируют многопериодную задачу определения местоположения зарядных станций в сети как задачу смешанного математического программирования и предлагают эвристический алгоритм ее решения. По их словам, эффективность этой модели и алгоритма обосновывается эталонными данными, относящимися к дорожной сети Sioux Falls, и оба результата будут

использованы в рамках исследования, предполагаемого быть проведенным для штата Южная Каролина.

В [229] проблема поиска места для размещения зарядных станций решается с использованием информации о шаблонах пользования электромобилями. Именно, эти шаблоны идентифицируются исходя из потребности транспортных средств в зарядке а) как функция размера батареи и уровень ее заряда (взятых из Национального обзора поездок домашних хозяйств по пробегу транспортных средств и шаблонов назначения), и б) по данным по зарядке электромобилей в существующих реальных инфраструктурах зарядных станций. Авторы используют оба эти источника информации для определения местоположений для зарядных станций с высокой вероятностью зарядки на них.

Задача стратегического планирования развития инфраструктуры станций быстрой зарядки аккумуляторных электрических транспортных средств рассматривается в [230], и математическая модель, лежащая в основе формализации этой задачи в виде задачи смешанного математического программирования, предлагается в этой публикации. Решение этой задачи определяет места расположения для новых зарядных станций и наиболее подходящее время для открытия их там. Предлагается некоторый общий алгоритм для получения решения этой задачи и авторы сообщают о некоторых результатах формулировки и решения этой задачи с использованием предложенных ими математической модели и алгоритма. Эти результаты были получены в рамках тематического исследования, проведенного в штате Калифорния, и авторы указывают на набор факторов, влияющих на принятие плановых решений.

В [231] предложена математическая модель для определения оптимальной инфраструктуры набора зарядных станций для аккумуляторных электромобилей на основе данных, относящихся к городу Wuban. Авторы используют программное обеспечение Matlab для решения задачи оптимизации, которая сформулирована на основе этой модели и решается с целью снижения уровня чрезмерной разрядки электромобилей. Они сообщают результаты расчетов,

определяющих количество и расположение зарядных станций, чтобы обеспечить эту скорость. как для медленных, так и для быстрых зарядных станций, что получается в результате решения этой задачи оптимизации.

В [232] предлагается тематическое исследование, связанное с выбором регионов в пределах географической зоны, в которых должны быть размещены зарядные станции для аккумуляторных электромобилей для удовлетворения спроса на зарядку электромобилей, работающих в качестве такси. Авторы использовали данные, относящиеся к 800 автомобилям такси, совершающим частые поездки между разными пунктами отправления и назначения, для расчета спроса на зарядку. Далее они использовали этот рассчитанный спрос в их тематическом исследовании, в котором они рассмотрели и решили проблему поиска максимально возможного покрытия спроса на электроэнергию за счет ограниченного количества зарядных станций, в том числе уже существующих. Авторы считают, что их модель должна использоваться только для идентификации районов, в которых должны быть расположены зарядные станции, а не точных мест в этих регионах, где зарядные станции должны быть открыты. Определение этой возможности возникает в результате работы системы поддержки принятия решений, основанной на их модели. Они полагают, что поиск именно этих мест будет предметом, так называемой пост-оптимизации, проводимой местными органами власти, которые должны учитывать определенные параметры, не учтенные при решении проблемы покрытия спроса. Авторы сообщают об использовании их системы поддержки принятия решений в исследовательском проекте, проведенном в городе Вена в 2014 году.

В [233], ограниченный диапазон аккумуляторов и недостаточно развитая инфраструктура для быстрой зарядки электромобилей рассматриваются в качестве основных технологических барьеров, влияющих на проникновение электромобилей в семейство широко используемых пассажирских машин. Авторы пытаются оценить рентабельность а) производства электромобилей с бо-

лее длинным периодом работы батарей, и б) доступной инфраструктуры быстрой зарядки автомобилей, рассматривая коммерческие легковые автомобили эксплуатируемые в Германии по соответствующим лицензиям. Они предлагают свои выводы о существующем статус-кво. В частности, они упоминают феномен, суть которого в том, что значительный процент таких автомобилей используется в реальной жизни, в то время как инфраструктура системы зарядных станций остается недостаточно развитой. Также они приходят к выводу, что более значительное проникновение электромобилей в коммерческий парк, вероятно, произойдет из-за наличия большего числа автомобилей с более длинным диапазоном заряда батарей.

Попытка рассмотреть зарядную инфраструктуру для электромобилей в городе в зависимости от взаимосвязанных решений частных инвесторов в этой инфраструктуре и водителей транспортных средств, чтобы инвестировать в эту инфраструктуру, рассматривается в [234]. Хотя эти две группы агентов связанные с этой инфраструктурой а) могут даже не взаимодействовать друг с другом, и б) иметь разные целевые функции (максимизировать прибыль для инвесторов и максимизировать функции полезности водителей, связанные с маршрутами их путешествий), их взаимосвязанные решения влияют на зарядную инфраструктуру. Кроме того, эта инфраструктура зависит как от мощности сети электроснабжения зарядных станций электричеством, так и от городского трафика. Авторы предлагают нелинейную математическую модель для отражения поведения агентов, имеющих дело с зарядной инфраструктурой, и сформулировать на его основе задачи нелинейной оптимизации. Они предлагают интерактивный метод решения этой задачи и анализируют его сходимость. Авторы считают, что результаты, представленные в статье, создают теоретическую основу для инвестиционного планирования в разработку инфраструктуры зарядных станций, и они сообщают о результатах использования их метода решения для разработки сетевой инфраструктуры зарядных станций города Sioux Falls в Южной Дакоте.

Важное направление развития инфраструктуры зарядных станций для электромобилей – это использование подзарядки аккумуляторных батарей на станциях обмена этих батарей. Влияние использования этих станций для подзарядки электромобилей их водителями по графику электроснабжения, предлагаемому электрическими сетями, изучаются в [235]. Предлагается ввести определенные временные ограничения для водителей электромобилей, определяющие режим так называемой оптимальной зарядки. Именно, этим водителям рекомендуется не заряжать их электромобили в определенные периоды времени в 24-часовом сегменте, включая самые загруженные часы для такси-стов и три ночных часа. Авторы предлагают оптимизационную модель для формулировки задачи минимизации затрат на тепловую генерацию электроэнергии в сети электроснабжения, обеспечивающих баланс силовой нагрузки. Они считают, что расчет оптимального режима зарядки и рекомендации водителям придерживаться этого режима, стимулирует использование возобновляемых источников энергии в электроснабжении. Однако, хотя график подачи электроэнергии в инфраструктуру зарядных станций и число электромобилей, использующих эту инфраструктуру, образуют некоторую взаимосвязанную систему, в статье рассматриваются проблемы улучшения работы только генерирующей части этой системы.

Проблема оптимального размещения станций замены аккумуляторов для электромобилей изучается в [236] применительно к нескольким логистическим корпорациям, которые используют такие транспортные средства. Эти корпорации заинтересованы в определении а) планов маршрутов для своих электромобилей и б) стратегий зарядки, которые соответствовали бы оптимальному размещению станций для замены. Предлагается математическая модель, лежащая в основе постановки рассматриваемой задачи как задачи целочисленного программирования. В этой модели накладываются определенные ограничения как на дальность движения, так и на грузоподъемность, тогда как ни оптимальное количество депо, ни оптимальное их размещение не ищутся. Авторы сообщают об использовании своей модели и результатах решения



сформулированной на ее основе задачи целочисленного программирования с помощью нескольких эвристических методов, которые они сравнивают для различных количеств клиентов. Кроме того, они предлагают экономический и экологический анализ различных рабочих сценариев для потенциальных пользователей своей модели.

Идея маршрутизации транспортных средств, находящихся в транспортной сети, к станции замены батарей, на которой полностью заряженный аккумулятор доступен на момент прибытия автомобиля рассматривается в [237]. Авторы предлагают метод, сочетающий онлайн-маршрутизацию и резервирования батарей для минимизации средней задержки для всех транспортных средств, находящихся в транспортной сети, что достигается путем решения динамической задачи, формализуемой с использованием Марковского процесса принятия решений. Они сообщают о результатах решения задачи отыскания минимума на основе варианта метода динамического программирования, и они используют данные, доступные для сети автомобильных дорог штата Аризона.

Математическая модель для формализации задачи нахождения структуры набора станций замены батарей для обслуживания аккумуляторных электробусов предлагается в [238]. Эта модель лежит в основе задачи смешанного математического программирования, которая формулируется, чтобы найти а) число станций и их расположение, и б) типы и число автобусов, которые будут обслуживаться совокупностью этих станций. Авторы раскладывают эту задачу на две подзадачи. Решение одной из этих подзадач определяет места расположения и мощности станций замены батарей, тогда как вторая подзадача определяет тип и количество батарей, которые можно менять на каждой станции. Авторы сообщают о результатах тестирования предложенного ими подхода к рассматриваемой проблеме на примере небольшой сети и обсуждают, какие параметры батарей являются ключевыми для определения конфигурации системы станций для замены батарей.

Задача определения мест размещения зарядных станций для электромобилей для обслуживания набора электромобилей такси в городе, а также числа зарядных устройств на каждой из размещаемых станций, исходя из схем проезда такси, отражающих жизненные шаблоны поведения клиентов, рассматривается в [239]. Авторами предлагается математическая модель для формализации задачи размещения зарядных станций как задачи целочисленного программирования. Целевая функция этой задачи описывает объем инвестиций в инфраструктуру, образуемую этими зарядными станциями, тогда как одно из ограничений задачи учитывает загруженность на зарядных станциях в форме очередей. Это ограничение помогает оценить как число водителей, которые отказываются стоять в очереди, чтобы зарядить свой электромобиль, и вероятность того, что водители будут заряжать свои электромобили в местах своего проживания. Авторы сообщают о результатах применения своей модели для расчета числа зарядных станций и их распределения применительно к потоку электромобилей такси в городе, и они обсуждают экономические преимущества, которые, по их мнению, демонстрирует использование модели.

Многокритериальная модель предложена в [240] для формулировки задачи математического программирования, которую предлагается решить, чтобы найти оптимальное расположение набора станций зарядки электромобилей и станций замены аккумуляторов. Идея состоит в том, чтобы найти такие места расположения обоих типов этих объектов, которые позволили бы электромобилю быстро заменить разряженные батареи на уже заряженные. Авторы сообщают о результатах тестирования своей модели и о методах, которые они использовали для решения многокритериальной задачи, сформулированной на основе этой модели, полученных ими для автобусной тестовой системы IEEE.

Оптимизационная модель для формализации и решения задачи размещения общественных зарядных станций для электромобилей, в которых целевая функция—это число миль, которое электромобиль преодолевает, оставаясь при этом заряженным, предложена в [241]. Авторы формулируют задачу как

задачу математического программирования, и они используют шаблоны движения транспортных средств, чтобы правильно отразить спрос на зарядку на таких зарядных станциях. Задача оптимизации сформулирована авторами с использованием Общей Системы Алгебраического Моделирования и результаты расчетов получены с использованием пакета CPLEX для решения оптимизационных задач. Авторы сообщают о применении своей модели для расчетов с реальными данными для траекторий перемещения такси в Пекине. В частности, они сообщают о результатах сравнения количества миль, которое электромобиль может проезжать оставаясь заряженным при оптимальном расположении зарядных станций в соответствии с рекомендациями их модели. с тем количеством миль, которые электромобиль покрывает при существующем местоположении зарядных станций.

Смысл проведенного обзора в том, чтобы убедиться, что постановка задачи создания, развития и управления функционированием городской сети зарядных станций для электромобилей идентична задаче создания, развития и управления функционированием сетью зарядных станций для электромобилей на скоростных автомобильных магистралях региона, которая была подробно изучена в [215]. Разумеется, математические модели в этих сетях отличаются за счет некоторых дополнительных ограничений, которые присутствуют в модели, используемой при формулировке задачи создания, развития и управления функционированием городской сети зарядных станций для электромобилей. Однако, структура системы ограничений в обеих моделях одинакова, и, полученные в [215] математические результаты могут быть использованы при формулировке и решении задачи для городской сети зарядных станций.

#### 4.2.3 Построение математической модели рассматриваемой задачи

Идея, лежащая в основе подхода к формализации задачи об оптимальном размещении зарядных станций для электромобилей в городе, рассматриваемой в настоящем исследовании, та же, что была предложена в [215] для задачи размещения таких зарядных станций вдоль скоростной автомагистрали, пересекающей некоторый географический регион и состоит в выборе

особой структуры математической модели для этой задачи. Такая структура позволяет формализовать задачу как задачу математического программирования, которая может быть эффективно решена с использованием стандартных пакетов программ. Структура модели такова, что позволяет оценить инвестиции, необходимые для развития сети зарядных станций для электромобилей в городе при неопределенности по а) экспертным оценкам потребности электромобилей в электроэнергии на каждой зарядной станции сети, и б) ценам на электричество, объявляемым электрической сетью (сетями), питающей (питающими) электричеством сеть этих зарядных станций.

Эффективность решения формализованной задачи в частности подразумевает возможность получения количественных оценок размера инвестиций даже когда а) эта проблема является крупномасштабной (т. е. имеет большое количество переменных и ограничений) и б) некоторые из переменных задачи могут быть целыми. Такую возможность обеспечивает предложенная в [215] математическая модель (с незначительными изменениями системы ограничений), лежащая в основе формализованного описания рассматриваемой задачи. К тому же, решение формализованной задачи, которая формулируется на основе предложенной модели, может быть найдено за короткий промежуток времени, и отыскание этого решения не требует разработки каких-либо новых математических методов и алгоритмов. То есть, это решение может быть найдено с использованием стандартных программных пакетов, позволяющих быстро и точно решать крупномасштабные оптимизационные задачи.

Два хорошо известных пакета программ

а) CPLEX (<https://www.ibm.com/analytics/cplexoptimizer>) и

б) MILP (<https://www.mathworks.com/help/optim/ug/intlinprog.html>)

являются примерами таких программных пакетов для решения задач математического программирования с непрерывными и дискретными переменными, включая задачи линейного, целочисленного и смешанного математического программирования.

Нелинейные функции присутствуют в предлагаемой математической модели. Их наличие связано с вышеупомянутой неопределенностью по экспертным оценкам потребности в электроэнергии и по экспертным оценкам цен на электроэнергию. Именно, эти оценки делаются, когда ни стоимость обслуживания конкретного электромобиля определенного типа на конкретной зарядной станции в конкретный период времени, ни числа электромобилей каждого типа для обслуживания на этой станции в этот период времени известны точно заранее. В результате математическая модель, которая будет использоваться для формализации рассматриваемой задачи, оказывается нелинейной.

При разработке математической модели для этой задачи, для большей определенности, сделано предположение относительно структуры спроса на электроэнергию. Именно, предполагается, что эта структура является одинаковой на каждой зарядной станции из сети в каждые 24 часа—с полуночи до следующей полуночи—в каждом из четырех времен года (зима, весна, лето и осень), хотя, как правило, эти структуры разные для разных времен года. Однако, как ясно из структуры предлагаемой математической модели (и как это отмечалось в [215]), предлагаемый подход к моделированию задачи не ограничен количеством «сезонов» внутри которых эти структуры спроса одинаковы и число таких «сезонов» может быть любым, от четырех до 365/366 ежегодно.

В рамках настоящего исследования рассматривается только случай, когда предполагается, что вся электроэнергия поступает на все зарядные станции в сети этих станций от электросети (или от разных электрических сетей на разные зарядные станции) напрямую. Однако, как упоминалось выше, существует подход к расширению использования предлагаемой математической модели, охватывающий, в частности случаи, когда электричество к зарядным станциям может также поставляться из а) электрической сети (или других электрических сетей) через системы хранения электроэнергии, и б) от возобновляемых источников энергии (опять же, напрямую и через системы хранения электроэнергии). Этот подход кратко описывается в [215].

Предположим, что задача построения и эксплуатации сети станций зарядки электромобилей в городе некоторого географического региона формулируется для заинтересованного юридического лица с целью оценки размеров необходимых для этого инвестиций.

Пусть, как и в [215], используемая математическая модель связывает:

- число точек в городе, в каждой из которых должна быть установлена зарядная станция, в соответствии с существующими федеральными и региональными правилами;

- число часовых сегментов времени в течение каждых 24 часов (24 сегментов);

- ожидаемый (средний) спрос на электроэнергию на каждой зарядной станции в каждый конкретный период времени;

- ожидаемые максимальные и минимальные потребности в электроэнергии на каждой зарядной станции в каждый конкретный период времени;

- ожидаемую (среднюю) розничную цену на электроэнергию (то есть цену, по которой электричество, как ожидается, будет продаваться клиентам каждой зарядной станции) в течение каждого конкретного временного сегмента;

- ожидаемую максимальную розничную цену на электроэнергию на каждой зарядной станции в каждый конкретный временной сегмент (которая продиктована ожидаемым статусом рынка и административными правилами) в любое время года;

- (среднюю) цену на электроэнергию, по которой, как ожидается, электроэнергия будет приобретена заинтересованным юридическим лицом для каждой зарядной станции из (соответствующей) региональной электрической сети в течение каждого конкретного периода времени;

- максимальную и минимальную цену на электроэнергию, по которой, как ожидается, заинтересованное юридическое лицо будет покупать электроэнергию для каждой зарядной станции из (соответствующей) региональной электрической сети в каждый конкретный период времени;

- объем электроэнергии, который заинтересованное юридическое лицо предполагает купить в соответствующей электрической сети для каждой зарядной станции в каждый конкретный период времени в любое время года;
- число типов электростанций (то есть количество типов этих станций), которые можно купить, установить и использовать на каждой зарядной станции из сети этих станций в городе;
- суточную мощность электростанции каждого конкретного типа на каждой зарядной станции;
- ожидаемое максимальное значение суточного объема электроэнергии (за каждые 24 часа), которое необходимо купить для каждой зарядной станции;
- число электростанций каждого типа, которое планируется установить на каждой зарядной станции;
- ожидаемые расходы, связанные с приобретением и установкой электростанций каждого типа на каждой зарядной станции из сети;
- ожидаемые ежедневные затраты на техническое обслуживание, связанные с обслуживанием электростанции каждого типа на каждой зарядной станции в каждое из времен года;
- ожидаемую стоимость эксплуатации каждой зарядной станции в течение определенного периода времени в каждом из времен года;
- количество дней в каждом сезоне четырех времен года (зима, весна, лето и осень);
- положительные вещественные числа, отражающие ожидаемые максимальные и минимальные уровни суточного (за каждые 24 часа) спроса на электроэнергию на каждой зарядной станции в каждый из четырех времен года;
- количество лет, на которые заинтересованное юридическое лицо оценивает объем инвестиций;
- уровень налога (процент от дохода), установленный для сети зарядных станций;

– положительное действительное число, отражающее возможные потери электроэнергии, которая, как ожидается, будет куплена в электрической сети (или в нескольких электрических сетях) и доставлены на зарядные станции, составляющие цепь, включая потери, связанные с передачей электроэнергии, купленной из электрической сети (сетей), для зарядных станций сети;

– ежедневный объем “краудфандинга” для каждой зарядной станции, который ожидается получить в каждый сезон.

Чтобы оценить необходимый объем инвестиций в разработку и эксплуатацию сети зарядных станций для электромобилей в городе, следует рассчитать:

– долю ежедневного дохода, ожидаемого к получению заинтересованным юридическим лицом в результате работы сети зарядных станций (т.е. каждые 24 часа в сутки) в каждое из четырех времен года, остающуюся после вычитания из этого дохода денежной суммы, подлежащей уплате в виде налогов;

– сумму денег, которая, как ожидается, будет ежедневно выплачиваться электрической сети (электрическим сетям) в каждом из четырех времен года за электроэнергию, которая будет куплена для зарядных станций, входящих в цепь;

– расходы, связанные с приобретением и установкой генераторов мощности всех типов на всех зарядных станциях, входящих в сеть;

– ежедневные расходы, связанные с обслуживанием всех генераторов мощности всех типов на всех зарядных станциях, входящих в сеть;

– ежедневные расходы, связанные с эксплуатацией сети зарядных станций в каждом из четырех времен года;

– сумму денег, которая, как ожидается, будет ежедневно собираться в каждом из четырех времен года от владельцев электромобилей и других членов сообществ, проживающих на территориях, близких к зарядным станциям сети.

Следует иметь в виду, что возможности электрогенераторов, которые будут приобретены и установлены на зарядные станции, должны позволять



достичь объемов соответствующих максимальным суточным потребностям в электроэнергии каждой зарядной станции, входящей в рассматриваемую их сеть. Следует также иметь в виду, что каждый генератор электроэнергии каждого типа, который устанавливается на каждой зарядной станции, входящей в сеть зарядных станций, занимает некоторую площадь. При этом общая площадь, которая может быть занята всеми электрогенераторами всех мощностей, размещаемыми на всех зарядных станциях, ограничена, а также, что каждому генератору в свою очередь требуется определенная площадь для одновременного предоставления услуг заряжаемым им электромобилям.

Нетрудно показать, что учет всех этих условий может осуществляться в рамках системы линейных неравенств, связывающих соответствующие переменные модели и параметры территорий, на которых предполагается размещать зарядные станции. Эта система может иметь и не иметь допустимого решения, что может быть установлено с помощью методики, аналогичной той, которая предложена в [189]. Эта же методика позволяет сформулировать задачу отыскания максимально физически возможного числа электрогенераторов, которые можно установить на каждой зарядной станции, как задачу математического программирования со смешанными переменными.

Одной из особенностей задачи размещения сети зарядных станций для электромобилей в городе по сравнению с задачей их размещения вдоль скоростной автомагистрали состоит в ограничениях на выбор мест в пределах городской черты, где такие станции могут быть установлены. Если на автомагистрали все, в основном, определяется потоком электромобилей в каждом из направлений и близостью густонаселенных районов к конкретным точкам на автомагистрали, то в городе, особенно крупном, существенное значение имеет городская инфраструктура, определяющая близость к детским яслям и садам, дошкольным и школьным учебным заведениям, лечебным учреждениям, учреждениям культуры (концертным залам, театрам) и др. Более того, городские власти могут указать те участки городской территории, где открытие заправок станций для электромобилей и какой максимальной мощности будет

разрешено и где принципиально запрещено. Тем самым, математическая модель для задачи оптимизации размещения зарядных станций для электромобилей в городе должна включать дополнительные ограничения, которые оказывается возможным представить в виде неравенств с линейными и билинейными функциями тех же переменных, которые используются в модели, предложенной в [215].

При математической формулировке рассматриваемой задачи принимаются некоторые естественные предположения об указанных выше параметрах и переменных, которые связываются математической моделью, лежащей в основе формулировки этой задачи.

1) Предполагается, что стоимость обслуживания остается неизменной и принимает те же самые значения внутри каждого временного сегмента в каждое из времен года, в то время как, как правило, это стоимость разная для разных временных сегментов в течение 24 часов для разных зарядных станций. Хотя это предположение может выглядеть как упрощение реальности, можно легко убедиться, что это не влияет на структуру функции, описывающей ожидаемую остающуюся прибыль, которая может быть получена заинтересованным юридическим лицом в результате управления сетью городских зарядных станций для электромобилей (после уплаты налогов) на один год. Это является следствием присутствия в структуре указанной функции компоненты, которая не зависит от того, как вычисляется ежедневная стоимость обслуживания. Кроме того, хотя эта стоимость обычно остается неизменной в течение одного времени года в каждом году (что соответствует подходу управляющих компаний к найму персонала для обеспечения обслуживания), эта стоимость может незначительно меняться (увеличиться) каждый год.

2) В общем случае можно считать, что ожидаемая суточная стоимость эксплуатации электрогенератора каждого типа на каждой зарядной станции является постоянной. Тогда суточная стоимость эксплуатации каждой зарядной станции может быть описана как функция векторной переменной точно так же, как описывается стоимость обслуживания зарядной станции. Однако в

рамках настоящего исследования (как и в [215]) предполагается, что эта стоимость разная в пределах каждого часа в пределах каждых 24 часов в разное время года для каждой зарядной станции.

3) С целью упрощения изложения предполагается, что развитие сети станций зарядки электромобилей в городе завершается в течение года, в котором оно началось. Таким образом, предполагается, что вся сеть зарядных станций может начать свою работу с самого начала года, следующего за годом, в котором начинается развитие сети. Однако этот период времени может быть длиннее или короче одного года. Если это так, то прибыль должна сначала рассчитываться за период времени, который заканчивается в конце года, в котором вся сеть начинает свою работу, а затем прибыль следует рассчитывать ежегодно.

4) Все перечисленные выше изменяющиеся параметры математической модели целесообразно рассматривать как компоненты некоторых семи векторных переменных имеющих следующий смысл:

- спрос на услуги городской сети зарядных станций для электромобилей;
- объем электричества, покупаемого у сетей электроснабжения региона, в котором расположен город, для снабжения зарядных станций для электромобилей, входящих в городскую сеть;
- число электрогенераторов разных типов, устанавливаемых на зарядных станциях, включенных в городскую сеть зарядных станций для электромобилей;
- спрос на услуги зарядных станций, входящих в городскую сеть для электромобилей;
- средняя рыночная цена на электроэнергию, покупаемую у сетей электроснабжения региона для зарядных станций городской сети;
- стоимость услуг по управлению зарядными станциями, входящими в городскую сеть зарядных станций для электромобилей;

– ожидаемый объем "crowdfunding" как одного из источников финансирования работы сети зарядных станций для электромобилей.

Ясно, что заинтересованное юридическое лицо управляет первыми тремя из указанных выше семи векторных переменных. Вообще говоря, в системе ограничений математической модели могут присутствовать ограничения, связывающие любую пару из этих трех переменных и даже все три. Однако, в рамках настоящего исследования предполагалось, что все эти три переменные изменяются независимо. Напротив, оставшиеся четыре векторные переменные должны рассматриваться как рыночные, так как ими нельзя управлять в принципе (хотя бы заинтересованному юридическому лицу, которое заинтересовано в разработке и эксплуатации сети городских зарядных станций для электромобилей). Более того, четыре рынка, которым эти четыре векторные переменные принадлежат разные.

Чтобы убедить потенциальных инвесторов вложить средства в развитие городской сети зарядных станций для электромобилей следует дать оценку ожидаемой прибыли, которую заинтересованное юридическое лицо может получить при "наиболее неблагоприятном сценарии." При таком сценарии наиболее неблагоприятное поведение рынка (т.е. такое, при котором а) спрос на услуги зарядных станций, б) цены на электричество, которые могут предложить сети электроснабжения в регионе, в) рыночные цены на оказание услуг по обслуживанию сети зарядных станций и г) объем средств, которые можно ожидать собрать в ходе "crowdfunding") должны приниматься во внимание.

С этой целью, предлагается:

а) рассматривать ожидаемую прибыль как функцию указанных выше семи векторных переменных;

б) рассматривать минимум этой функции по тем четырем векторным переменным (на Декартовом произведении множеств ожидаемых значений каждого из этих переменных), которые являются рыночными и не зависят от решений, принимаемых заинтересованным юридическим лицом, и максимизи-

ровать этот минимум по тем трем векторным переменным (на Декартовом произведении множеств допустимых значений каждого из этих переменных), которые контролируются заинтересованным юридическим лицом. Эта функция имеет структуру, аналогичную структуре функции, рассмотренной в [215], где доказано, что при тех же предположениях о структуре системы ограничений для задачи, рассматриваемой в настоящем исследовании, которые выполнялись для задачи, рассмотренной в [215], а) функция минимума определена в каждой точке указанного выше Декартова произведения четырех множеств и б) максимум этой функции минимума достигается на указанном выше Декартовом произведении трех множеств.

Таким образом, в этом разделе отчета приведены результаты исследования моделей по оптимальному размещению объектов и привлечению частных инвестиций в рамках теоретико-игрового подхода. В частности:

- при решении задачи оценивания экономической целесообразности создания нового грузового транспортного узла, сформулировано и доказано достаточное условие конкурентоспособности тарифов на услуги, в таком узле, сформулированы все рассматриваемые в нем прикладные задачи и доказаны соответствующие утверждения;

- продемонстрирована возможность формулировки задачи об оценке финансовой целесообразности строительства и эксплуатации нового транспортного узла в виде задачи нахождения максимума максимина разности двух билинейных функций; показана возможность отыскания ее решения из решения вспомогательной задачи квадратичного программирования.

- предложена модификация ранее разработанной модели оптимального размещения зарядных станций автомобилей вдоль участка шоссе для решения аналогичной задачи размещения городских сетей таких станций.

## **5 Анализ данных и принятие решений в задачах социально-экономического развития регионов**

В этом разделе отчета отражены результаты по применению методов анализа данных к решению ряда прикладных задач социально-экономического развития регионов, а именно:

- исследование миграции в арктическом регионе;
- раздел спорных территорий арктического региона;
- анализ уязвимости социально-экономического развития Арктики из-за изменения климата;
- оценка водообеспеченности стран Африки;
- анализ продовольственной безопасности стран Африки и её зависимости от изменения климата;
- применение интервального оболочечного анализа данных к оцениванию эффективности превентивных противопожарных мероприятий.

### **5.1 Миграция в арктических регионах**

#### **5.1.1 Общий анализ проблемы миграции в арктическом регионе**

Арктика – это уникальная местность с впечатляющим, но очень хрупким биоразнообразием. Огромный потенциал Арктики заключается в том, что там находится много природных ресурсов от природного газа и нефти до меди, железа и золота, 22% неоткрытых, но технически извлекаемых ресурсов всего мира, 90 миллиардов баррелей нефти, 44 миллиарда баррелей сжиженного природного газа [242]. Коренные народы проживают на этой территории с давних времен. Глобальные экономические, климатические и экологические изменения сильно влияют на арктическую экосистему и население. И это имеет заметный и ощутимый след для Земли в целом. Арктика стала темой глобального интереса из-за новых изменений и ресурсного потенциала.

Экономические и межправительственные проблемы постоянно возникают во всем мире, заметно меняя обычный образ жизни людей за последние несколько лет. В 2015 году миграционный кризис стал основной проблемой

Европейского Союза. Исторически миграция была обычным явлением. С середины 1980-х годов в Северной Европе было довольно много мигрантов, но тем не менее, приток сотен тысяч мигрантов в поисках убежища был необычным явлением для Северных стран, которого никогда до этого не происходило. Большинство беженцев решили поселиться в странах Северной Европы, поэтому Арктическому региону нужно было найти решения новых сложных предстоящих проблем, помимо тех, которые там уже были. Сообщества коренных народов Арктики также сталкиваются с огромными миграционными изменениями, которые происходят во всем мире. Саамы в Скандинавии, а также коренные народы из России и Аляски (США) покидают свою естественную среду обитания и теряют свою историческую идентичность. Многие российские арктические города становятся заброшенными, и там остаются только пожилые люди и вахтовики [243]. Для решения сегодняшних проблем миграции необходимы партнерство и сотрудничество [244].

Миграция в арктическом регионе - очень важная и актуальная тема, но в то же время достаточно сложная и противоречивая. Миграция на территории Арктики актуальна для экономик многих стран. В данном исследовании обсуждаются различные стороны миграции, приводятся статистические данные, собранные из различных источников, также показываются различные формы культурных, экологических и экономических проблем населения Арктики. Очень важно понять размер, состав и темпы роста населения Арктики для прогнозирования будущего населения Арктики, что в свою очередь является полезным инструментом для политиков, позволяющий оценить, отличается ли ожидаемая демографическая ситуация в их регионе от желаемой.

Ключевой вопрос данного исследования: каково будущее населения Арктики?

Цели исследования:

- продемонстрировать важность миграции для устойчивого развития в мире;
- представить текущую структуру населения в Арктике;

- проанализировать миграционные тенденции в Арктике;
- спрогнозировать будущие тренды миграции в арктических регионах.

### 5.1.2 Анализ данных по миграции в арктическом регионе

Данные о глобальных миграционных потоках весьма ограничены. База данных о миграции содержит информацию только о 45 странах [245].

В этом исследовании используются статистические данные и факты из следующих организаций.

- IOM (international organization of migration);
- ArcticStat;
- Nordregio;
- UN population division;
- Demoscope;
- National statistical offices (U.S. Census Data and Statistics, Statistics Canada, Statistics Greenland, Statistics Faroe Islands, Statistics Iceland, Statistics Norway, Statistics Sweden, Statistics Finland, Rosstat).

На территории Арктики проживает более 5 миллионов жителей. Размер населения в разных частях Арктики сильно отличается: арктические регионы имеют разные уровни городской застройки, численность коренных народов и уровень рождаемости, смертности и переселения [246].

Основной официальной экономической деятельностью в Арктике является добыча полезных ископаемых, которые имеют решающее значение для экономического роста арктических государств. Однако жители Арктики в значительной степени полагаются на государственные субсидии, потому что добыча полезных ископаемых в Арктике обычно принадлежит сторонним организациям, не базирующимся на Севере. Здесь также присутствует малый локальный бизнес (охота, организация туров, ремесла), который помогает жителям Арктики зарабатывать себе на жизнь [247].

Рассмотрим последние демографические тенденции в Арктике, а также прогнозы на будущее. Область нашего исследования охватывает поселения,



которые живут в естественной среде обитания или имеют характеристики, аналогичные коренным народам Арктики.

Изменение населения в Арктике из-за естественного прироста и чистой миграции населения отличается для разных штатов или районов и состоит из двух компонентов:

- чистая миграция – разница между людьми, мигрирующими в регион, и теми, кто уезжает;
- естественный прирост – разница между количеством рождений и смертей.

Население мира с 1990 г. увеличилось на 43,2% (с 5,3 до 7,6 млрд человек) из-за естественного прироста [248]. В Таблице Б.1 Приложения Б представлены данные изменения численности населения с 1990 по 2018 год в 8 арктических государствах (США, Канада, Дания, Исландия, Норвегия, Финляндия и Россия).

На рисунке 5.1 показаны данные по продолжительности жизни в арктических регионах.

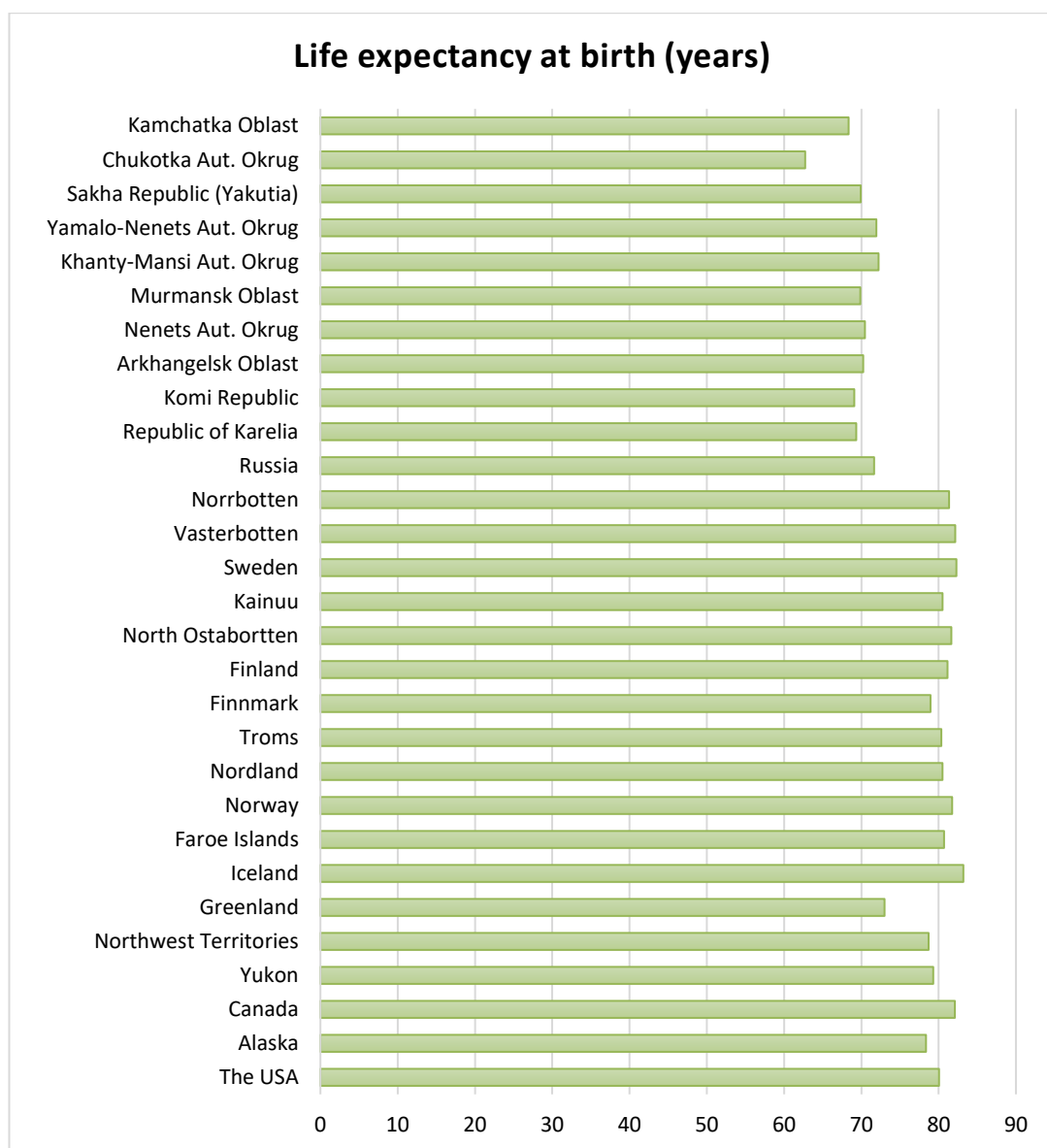


Рисунок 5.1 – Ожидаемая продолжительность жизни при рождении (лет) в арктических регионах, 2014–2019 гг. Источники: Statista, Statistics Finland, Statistics Norway, Rosstat, Statistics Canada, 2014-201

Все представленные в таблице арктические регионы можно разделить на 4 группы в зависимости от продолжительности жизни.

Группа 1. Исландия, Фарерские острова, арктические территории Норвегия и Швеция. Ожидаемая продолжительность жизни в этих странах высока: около 80 лет для мужчин и 84 для женщин. Средняя продолжительность жизни в этих регионах одна из самых высоких в мире.

Группа 2. Аляска, Юкон, Северо-Западные территории и Арктические территории Финляндии. Ожидаемая продолжительность жизни немного ниже, чем в скандинавских арктических регионах.

Группа 3. Гренландия. Средняя продолжительность жизни составляет 75 лет для женщин и 70 лет для мужчин.

Группа 4. Северные регионы России. Продолжительность жизни в этих регионах намного ниже, чем в среднем по стране. Средняя продолжительность жизни мужчин в арктических регионах составляет примерно 63 года, что меньше, чем в других странах. Средняя продолжительность жизни женщин на Русском Севере составляет около 74 лет. У женщин в России ожидаемая продолжительность жизни намного выше, чем у мужчин. Во всех северных регионах России ожидаемая продолжительность жизни ниже средней по стране, за исключением Ханты-Мансийского и Ямало-Ненецкого регионов. Средняя продолжительность жизни мужчин на Чукотке остается ниже 60 лет, что на 8 лет меньше общероссийского уровня.

Другой показатель – это переход к более низкой рождаемости. В большинстве арктических стран и регионов низкие показатели рождаемости. Это происходит из-за несвоевременного вступления в брак и рождения детей или увеличения бездетности. Данный показатель измеряют с помощью общего коэффициента фертильности, который является гипотетическим: это то, сколько детей, вероятно, родила бы женщина, если бы она забеременела в определенном возрасте (Таблица 5.1).

Таблица 5.1 – Общий коэффициент рождаемости в арктических регионах, 2014-2019 гг. Источники: Statista, Statistics Finland, Statistics Norway, Statistics Sweden, Rosstat, Statistics Canada, 2014-2019

Total fertility rate in selected Arctic regions (2014 - 2018)	
Region	Fertility rate (children born per woman)
The USA	1,73
Alaska	1,97
Canada	1,61
Yukon	1,73
Northwest Territories	1,97
Nunavut	2,97
Greenland	1,97
Iceland	1,99

## Продолжение Таблицы 5.1

Faroe Islands	2,34
Norway	1,85
Nordland	1,5
Troms	1,45
Finnmark	1,56
Finland	1,57
North Ostabortten	1,91
Kainuu	1,79
Sweden	1,71
Vasterbotten	1,68
Norrbotten	1,61
Russia	1,58
Republic of Karelia	1,52
Komi Republic	1,63
Arkhangelsk Oblast	1,56
Nenets Aut. Okrug	2,24
Murmansk Oblast	1,52
Khanty-Mansi Aut. Okrug	1,87
Yamalo-Nenets Aut. Okrug	1,9
Sakha Republic (Yakutia)	1,85
Chukotka Aut. Okrug	2,02
Kamchatka Oblast	1,65

С учетом некоторой смертности показатель 2,1 ребенка на женщину – это уровень, при котором население просто восполнит себя в долгосрочной перспективе. Отклонения от этого уровня имеют большое влияние на рост населения.

Коэффициенты рождаемости почти во всех арктических странах и регионах снижались за последние несколько десятилетий. Арктические регионы с высокой долей коренного населения, такие как Нунавут, Гренландия, Ненецкий округ и Чукотка, как правило, имеют более высокие коэффициенты рождаемости. В арктических регионах Норвегии (Нордланд, Тромс и Финнмарк) уровень рождаемости невероятно низкий из-за тенденции откладывать роды. Россия известна своим «демографическим кризисом», начавшимся в середине 1990-х годов и охватывающим также северные территории. Средний коэффи-

циент рождаемости в России довольно низкий, в Республике Карелия, Мурманской области и Архангельской области уровень рождаемости даже ниже среднего по России.

Демографические тенденции, вероятное будущее распределение и размер населения Арктики необходимы для государственного планирования, поскольку правительства могут уменьшить некоторые негативные последствия социально-экономических изменений. Все арктические государства регулярно делают прогнозы относительно своего будущего населения на региональном и национальном уровнях [249] (рисунок 5.2).

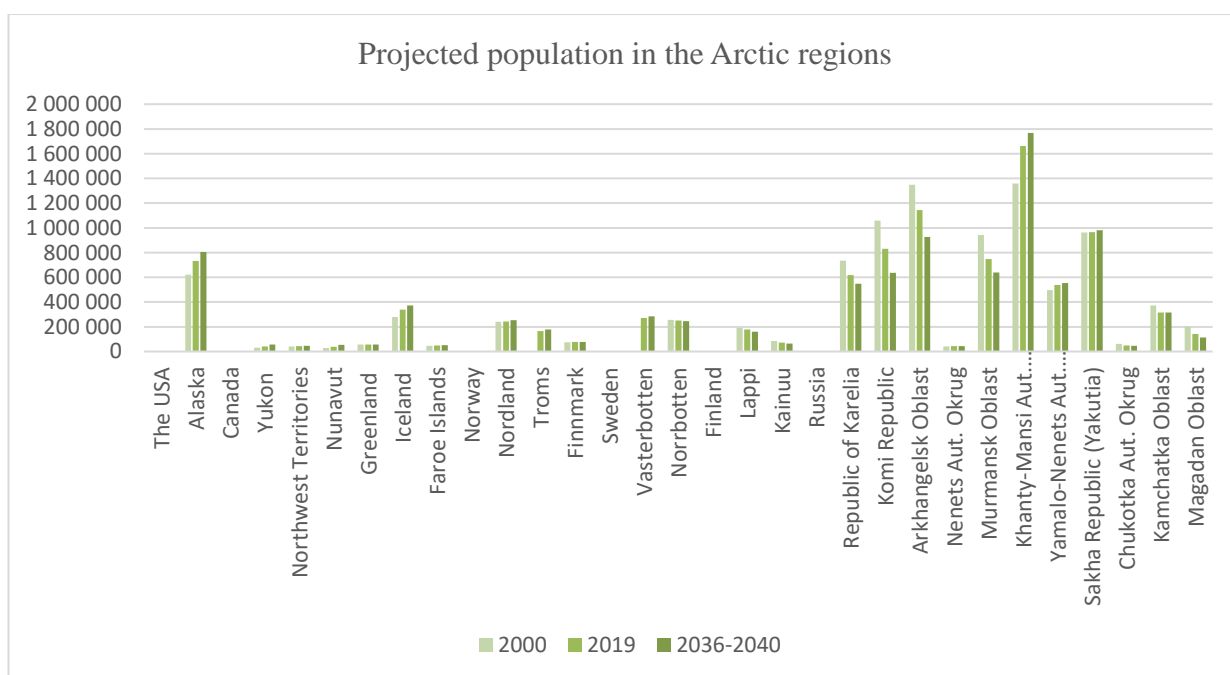


Рисунок 5.2 – Прогнозируемая численность населения в арктических регионах. Источники: Statista, Statistics Finland, Statistics Norway, Statistics Sweden, Rosstat, Statistics Canada, 2000-2020

Население мира будет продолжать расти. Эта тенденция особенно актуальна для менее развитых регионов. По оценкам ООН, население мира достигнет 9 миллиардов в 2037 году и 10 миллиардов в 2055 году. В прошлом рост мирового населения уже имел далеко идущие последствия для Арктики и продолжится в будущем.

В целом есть некоторые демографические тенденции, которые, по прогнозам, будут общими и будут иметь место в Арктике [250].

Во-первых, оно стареет: согласно прогнозам, в будущем население Арктики будет стареть, как и везде в мире. По прогнозам, относительный процент пожилых людей и пенсионеров будет расти во всех арктических регионах.

Во-вторых, соотношение полов: в изначально арктических поселениях было больше мужчин, чем женщин. Например, на Аляске в 2010 году было 108,5 мужчин на 100 женщин, а в 2045 году это соотношение снизится до 104,8 мужчин. В 1977 году в Гренландии было 118 мужчин на 100 женщин. Сейчас это соотношение составляет 112 мужчин на 100 женщин, и в будущем оно снизится.

И, наконец, урбанизация: очевидна тенденция переезда людей из небольших поселков в городские центры. Это происходит из-за необходимости получать качественное образование и перспектив трудоустройства в крупных городах. Итак, есть прогноз, что численность населения во многих небольших арктических поселениях сократится, крупные поселения будут расти.

Все эти тенденции важны для представления будущей структуры населения и должны быть приняты во внимание официальными лицами Арктики, чтобы иметь возможность адаптироваться к новым реалиям и планировать их на региональном и национальном уровнях.

COVID-19 сильно поменял жизни и экономическое благополучие людей в Арктическом регионе. Сильные изменения окружающей среды и климата усугубляют проблемы, с которыми сталкивается арктическое сообщество. Из-за коронавируса были отменены исследовательские экспедиции, туристическая деятельность и многие научные конференции. Часть коренных народов Севера испытывают жилищные и психологические проблемы. Многие домашние хозяйства не имеют свободного доступа к питьевой воде, кроме того, существуют проблемы с трубопроводами и транспортными сетями [251].

В целом последствия коронавируса выявили наиболее уязвимые проблемы коренных народов Арктики, а также привели к временной остановке экономической деятельности.

## **5.2 Распределение спорных территорий арктического региона**

В данном подразделе отчета отражены результаты по применению методов математического моделирования и анализа данных к распределению спорных территорий арктического региона.

### **5.2.1 Общий анализ проблемы спорных территорий арктического региона**

Активные исследования и добыча ресурсов в арктическом регионе начались относительно недавно. Причиной тому экстремальные климатические условия и труднодоступность этого региона. Однако одновременно с развитием технологий возрастал и спрос на энергию, кроме того, в Арктике изменялся климат, таяли ледники. Всё это привело к значительному росту заинтересованности стран в освоении и изучении арктического региона.

Арктика занимает около 6% поверхности земного шара. При этом, по некоторым оценкам, на её территориях находится примерно 30% потенциальных запасов газа и 13% потенциальных запасов нефти [252] со всей Земли. Около 84% этих ресурсов располагаются на не принадлежащих ни одной стране оффшорных континентальных шельфах [253].

Однако самое большое затруднение состоит в том, что активная эксплуатация нефти и газа негативно влияет экосистему. Загрязнение региона и возможные разливы нефти являются важным фактором, который обязательно нужно учитывать при принятии решения о том, как именно использовать арктический регион, так как в нём насчитывается около 450 видов рыб [254]. Важно понимать, что Арктика очень чувствительный регион. Это подчёркивает важность сотрудничества стран, присутствующих там и использующих арктический регион.

Всего существует 5 арктических государств: Россия, США, Канада, Дания и Норвегия. На территории Арктики все воды на расстоянии 200 миль (примерно 370 километров) от суши любого государства по закону являются его Исклyчительной Экономической Зоной (ИЭЗ), на эту зону у государства

суверенные права. Остальные воды, за этими пределами, считаются “открытым морем” и не принадлежат никому. Для регулирования споров, касающихся арктического региона, существует Арктический совет, отражающий международные интересы. В него входят: Канада, Дания, Финляндия, Исландия, Норвегия, Россия, Швеция и США. Также в качестве наблюдателей в нём состоят следующие страны: Великобритания, Германия, Испания, Италия, Китай, Республика Корея, Нидерланды, Польша, Франция, Япония, Индия и Сингапур. Одной из задач Арктического совета является рассмотрение претензий стран на арктические территории. Таким образом поднимается вопрос о том, каким именно способом наиболее справедливо распределить свободные арктические территории между заинтересованными странами.

Построение математической модели арктического региона с последующим поиском оптимальных методов распределения территорий, возможно, является одним из способов решения этого вопроса.

Основными задачами данной работы были определены:

- исследование уже имеющихся методов распределения спорных территорий арктического региона;
- предложение и проверка новых методов, способных улучшить результат уже существующих;
- программная реализация рассмотренных методов, применяемая не только к задаче арктического региона, но и любой другой территориальной задаче подобного вида;
- тестирование программной реализации на реальных данных задачи арктического региона и анализ полученных результатов.

#### 5.2.2 Анализ методов справедливого распределения спорных территорий и их модификация

В данной исследовательской работе использовалась постановка задачи из [255]. На основе данных из [256], лежащая севернее  $63^\circ$  северной широты арктическая территория делится на 640 000 равных участков, примерно по  $50\text{ км}^2$  каждый (рисунок 5.3). Помимо Арктических вод на данной территории



находятся территории 6 государств: США, Россия, Канада, Гренландия (Дания), Норвегия и Исландия.



Рисунок 5.3 – Карта (более светлыми тонами обозначена ИЭЗ)

Обозначим через  $X$  множество свободных территорий Арктики, а через  $Y$  множество стран, среди которых мы и будем их распределять. При учёте ИЭЗ более 12.5% участков остаются свободными, а это значит, что их можно распределить между указанными странами. Далее на карту добавляется информация о распределении следующих ресурсов: нефть, газ, рыба и морские пути (рисунок 5.4). Объём этих ресурсов на каждом участке обозначается за  $f(O, x)$ ,  $f(G, x)$ ,  $f(F, x)$  и  $f(M, x)$  соответственно. Информация об их расположении взята из [252], [257].

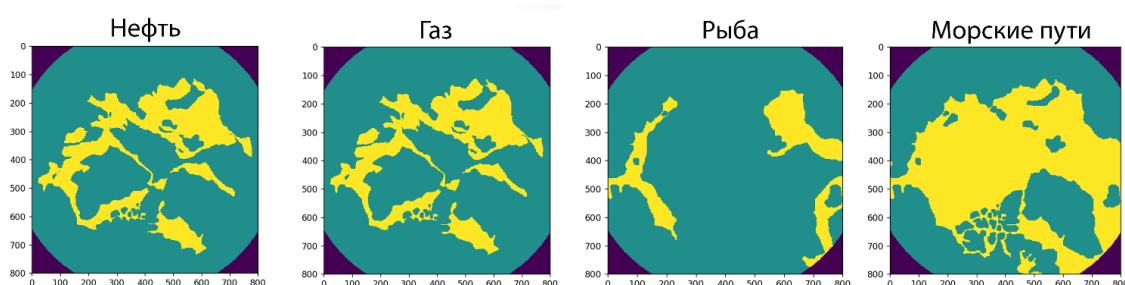


Рисунок 5.4 – Ресурсы (Фиолетовым обозначены не рассматриваемые территории, бирюзовым территории без ресурса, жёлтым – территории с ресурсом)

Последними данными, которые нужно получить для конечной формулировки задачи, являются уровни заинтересованности стран в ресурсах  $g(O, y)$ ,

$g(G, y)$ ,  $g(F, y)$  и  $g(M, y)$ . Однако по причине того, что никакой информации касательно уровней заинтересованности стран в ресурсах в открытых источниках найдено не было, в данной работе был предложен метод их оценки, основанный на соотношении количества поисковых выдач по запросу пары соответствующей страны и ресурса. Результат применения данной оценки показан ниже (Таблица 5.2).

Таблица 5.2 – Уровни заинтересованности стран в ресурсах

	Нефть	Газ	Рыба	Морские пути
США	4	7	4	1
Россия	5	6	3	2
Канада	5	6	4	1
Гренландия	4	6	5	1
Норвегия	4	6	4	2
Исландия	5	7	3	1

Задача состоит в распределении всех участков между странами так, чтобы уровни удовлетворённости  $S(y)$  одновременно всех стран были как можно больше. Такую целевую функцию можно обозначить как максимизацию минимального уровня удовлетворённости среди всех стран [255].

$$\min_y(S(y)) \rightarrow \max.$$

Идея расчёта уровня удовлетворённости страны, итоговым распределением территорий, следующая:

Заинтересованность  $s(x, y)$  страны  $y$  в участке  $x$  прямопропорциональна векторному произведению вектору уровней заинтересованности страны в ресурсах участка на вектор объёмов ресурсов этого участка и обратно пропорциональна коэффициенту минимального расстояния между ними.

$$s(x, y) = a(x, y) f(x) g(y),$$

$$f(x) g(y) = f(O, x) g(O, y) + f(G, x) g(G, y) + f(F, x) g(F, y) + f(M, x) g(M, y),$$

$$a(x, y) = \max(\min(d_2 - d_1, d_2 - h(x, y)), 0) / (d_2 - d_1),$$

где  $d_1$  и  $d_2$  – это произвольные границы расстояния. В [255]  $d_1$  принята за расстояние исключительной экономической зоны, то есть 370км, а  $d_2$  за 4000км.

Итоговый уровень удовлетворённости страны в предыдущих работах рассчитывался как разность суммы её заинтересованности в принадлежащих ей участках и суммы её заинтересованности в не принадлежащих ей участках.

$$S(y) = \sum_{x \in Y} s(x, y) - \sum_{x \notin Y} s(x, y).$$

Однако в данной работе предложена новая формула:

$$S(y) = \sum_{x \in Y} s(x, y) / \sum_{x \in X} s(x, y).$$

В таком случае, если считать передачу участков непрерывной, то можно доказать наличие решения, при котором уровни удовлетворённости всех участков будут одинаковы. Докажем с помощью математической индукции.

Рассмотрим случай с двумя участниками А и В. В качестве начального распределения отдадим все территории участнику А. Затем определим любую непрерывную функцию передачи территорий от А к В (рисунок 5.5).

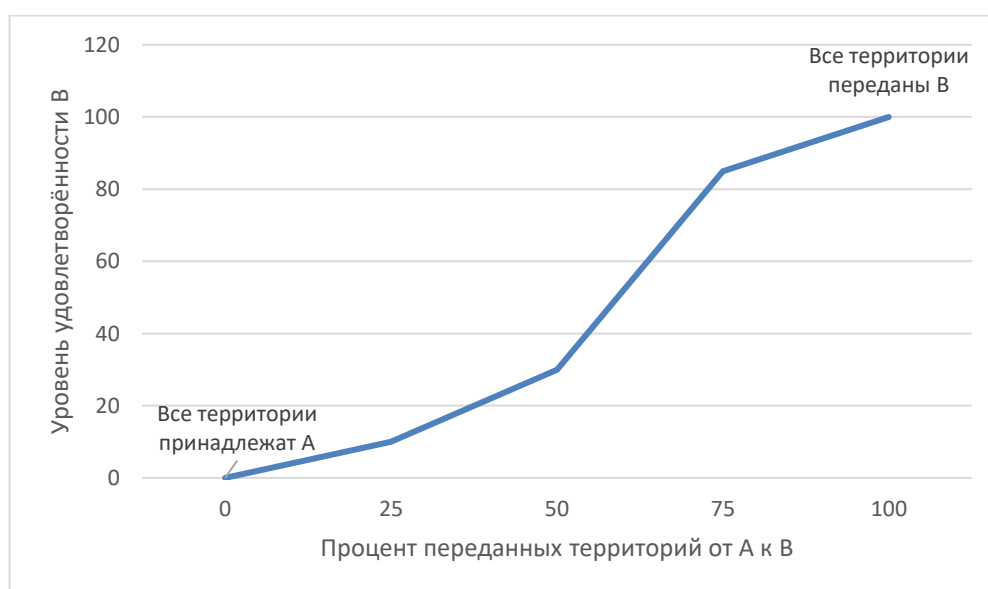


Рисунок 5.5 – График уровня удовлетворённости В по мере передачи территорий от А к В

Рассмотрим график разности уровней удовлетворённости участника А и В по мере передачи территорий (рисунок 5.6).

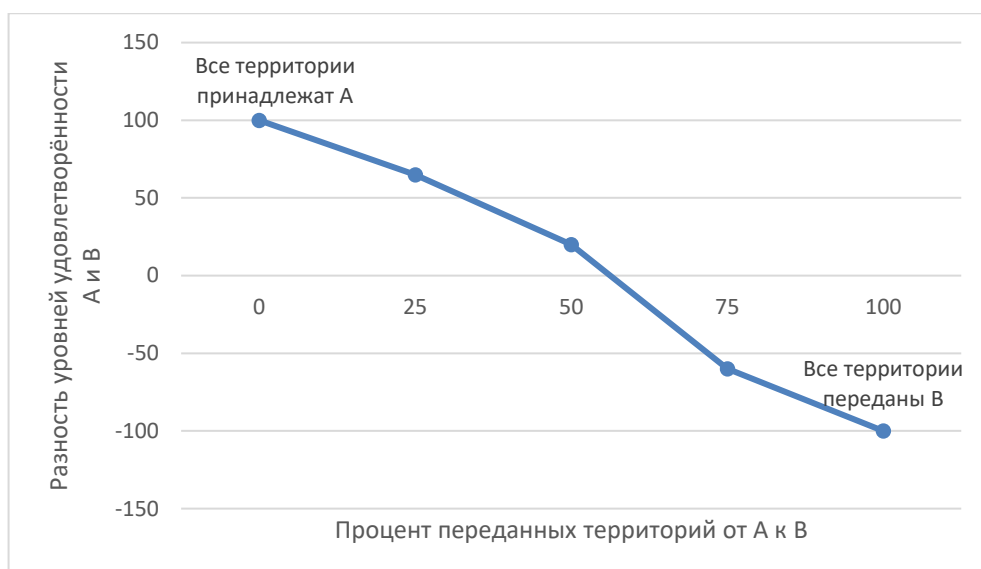


Рисунок 5.6 – График разности уровней удовлетворённости A и B по мере передачи территорий от A к B

На рисунке 5.6 заранее известны две точки: начало функции (разность равна 100, так как все территории принадлежат A), и конец функции (разность равна -100, так как все территории принадлежат B). В точке, где достигается равенство уровней удовлетворённости, значение их разности будет равно нулю, то есть пересекать ось абсцисс. Ввиду того, что начало функции имеет положительное значение, а конец – отрицательное, наличие такой точки гарантируется Теоремой Коши о нуле непрерывной функции.

Допустим, теперь у нас есть  $n$  участников, между которыми уже распределены все участки таким образом, что все уровни удовлетворённости участников равны  $k$ . Тогда, при добавлении нового участника D, нужно снова определить непрерывные функции передачи участков от каждого участника новому. Взяв сумму этих функций, получим функцию одновременной передачи участков от всех участников новому участнику D (рисунок 5.7).

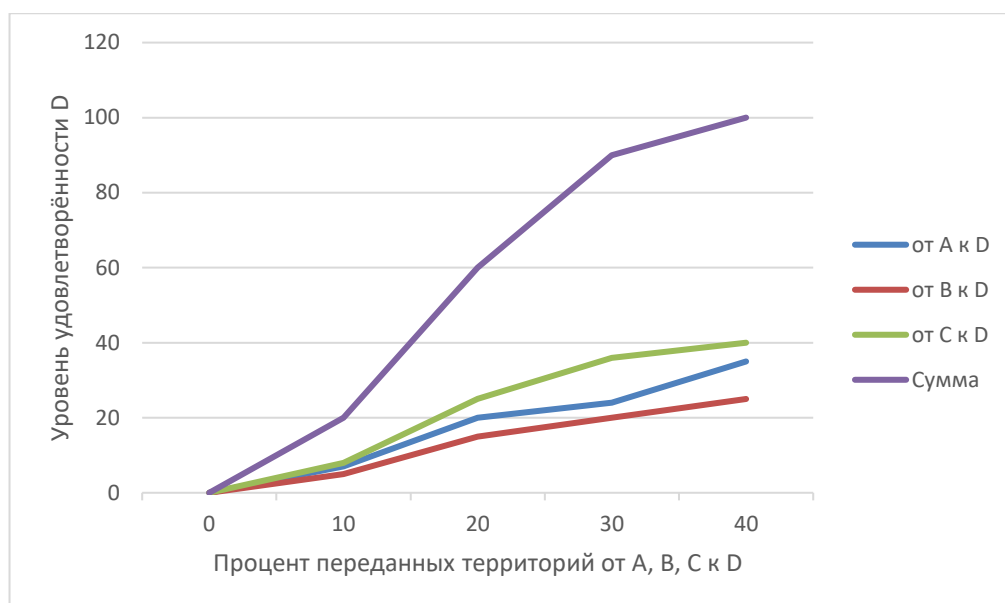


Рисунок 5.7 – График уровня удовлетворённости D по мере передачи территорий от A, B, C

Используя эту функцию, мы можем снова рассмотреть график разности уровней удовлетворённости старых участников и нового D (рисунок 5.8).

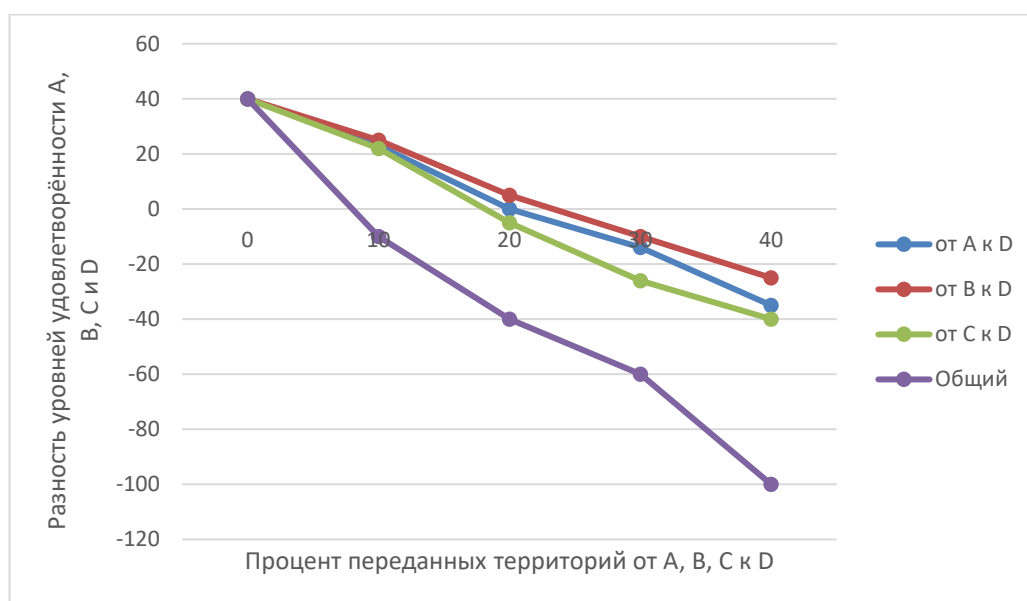


Рисунок 5.8 – График разности уровней удовлетворённости A, B, C и D по мере передачи территорий к D

В данном случае, началом функции будет точка k, так как уровень удовлетворённости старых участников k, а нового 0. Конечной точкой будет -100, так как в этот момент все участнику отдадут все свои территории участнику D, и их уровень удовлетворённости будет равен 0, а его 100. Ввиду того, что

начальный уровень удовлетворённости участников  $k$  не может быть отрицательным, мы можем снова применить Теорему Коши о нуле непрерывной функции и найти точку, в которой разность каждого участника и участника  $D$  будет равна нулю, а значит, все уровни удовлетворённости будут равны.

Теперь рассмотрим сами методы распределения территорий:

- метод максимальной пользы [255]: каждый свободный участок передаётся той стране, которой он принесёт больше пользы;

- метод минимального расстояния [255]: каждый свободный участок передаётся той стране, минимальное расстояние до которой меньше;

- метод благотворительности [255]: после распределения всех свободных участков те страны  $i$ , у которых уровень удовлетворённости больше, передают часть полученных участков странам  $j$  с меньшим уровнем удовлетворённости; при этом участки для обмена выбираются по правилу максимизации отношения  $s(x, y_j) / s(x, y_i)$ ;

- метод баланса, который был предложен в настоящем исследовании: страны в порядке увеличения уровня удовлетворённости выбирают самый интересующих их участок; эта процедура повторяется, пока не закончатся свободные участки;

- метод взаимовыгодного обмена, который был предложен в настоящем исследовании: после распределения всех свободных участков поочерёдно выбираются пары стран и с помощью задачи смешанного целочисленного линейного программирования находится такой обмен участками между этими странами, который выгоден им обоим.

В итоговом варианте программной реализации присутствует возможность моделирования территориальной задачи с любым числом стран, ресурсов, масштабом карты. Кроме того, имеется возможность выбора использовать относительный уровень удовлетворённости или стандартный, также присутствует возможность учёта сферической кривизны поверхности. Все вышеперечисленные методы распределения территорий также доступны. Присутствует возможность проанализировать результат с помощью различных карт,

анимаций, трёхмерных моделей (рисунок 5.9) и гистограмм. Ввиду того, что процесс расчётов долгий, имеется возможность сохранять и загружать модель.

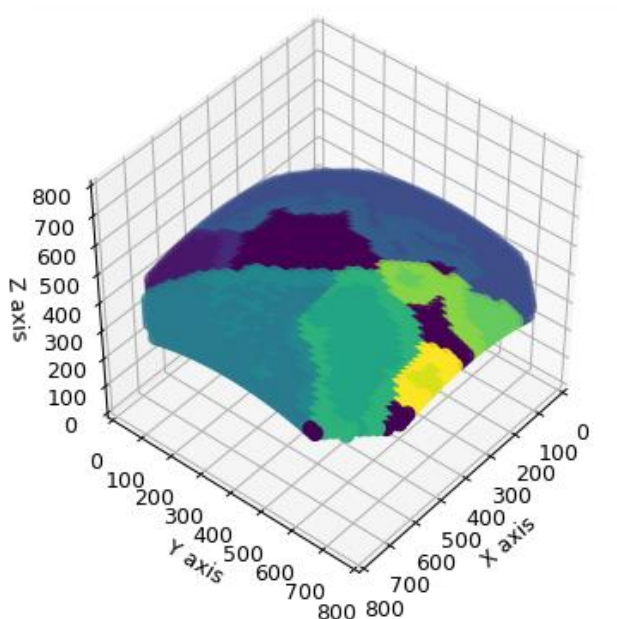


Рисунок 5.9 – Пример трёхмерного отображения заданной территории на сфере

Все вышеперечисленное реализовано на языке Python в среде разработки Jupyter Notebook. Код доступен на GitHub по ссылке <https://github.com/vampiirre/territorial-conflict/blob/master/model.py>.

При применении данной программной реализации методов на данных задачи арктического региона, были получены следующие результаты (рисунок 5.10).

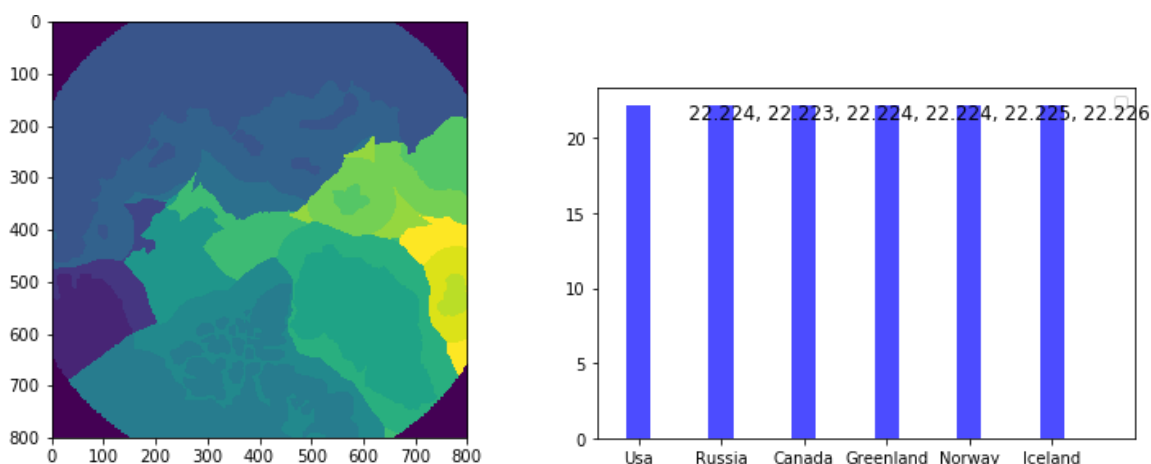


Рисунок 5.10 – Результат работы модели: итоговая карта территорий и гистограмма уровней удовлетворённости стран

Важно отметить, что все страны имеют доступ к своим участкам, кроме того, их уровень удовлетворённости примерно одинаков.

Распределение ресурсов по странам в процентном соотношении получилось следующее (рисунок 5.11).

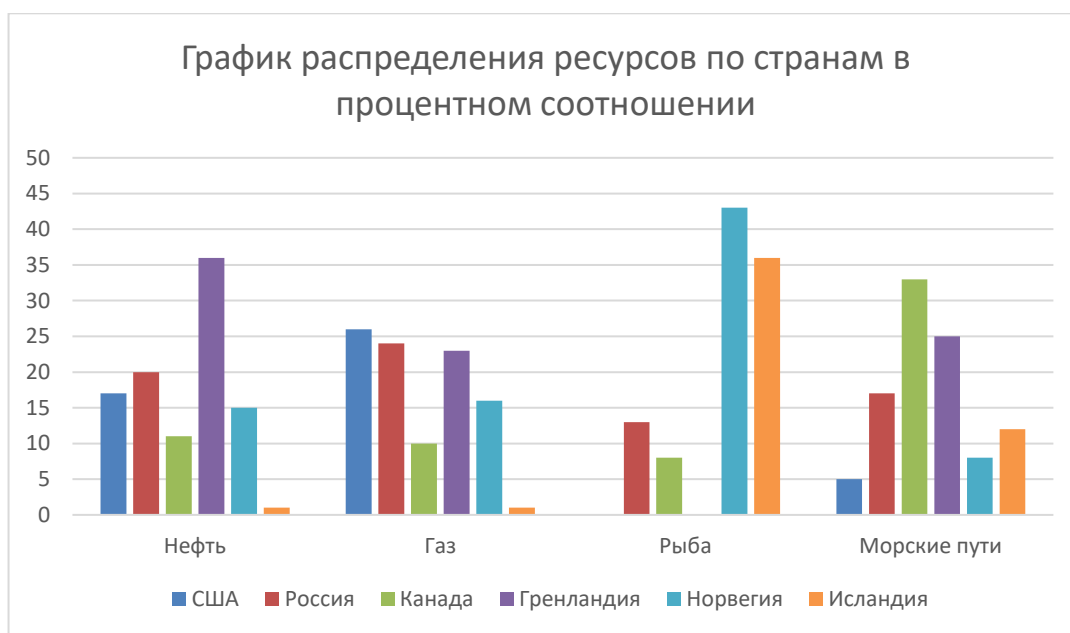


Рисунок 5.11 – График распределения ресурсов по странам в процентном соотношении

Всего было произведено более 457 тысяч итераций обмена.

Важное для рассмотрения свойство – свобода от зависти, оно говорит о том, желает ли участник А вместо полученных территорий получить территории участника Б. Значение зависти равно разнице уровней удовлетворённости



участника от его территорий и от чужой, если оно отрицательно, то считается, что участник завидует другому.

Таблица 5.3 показывает, что имеется 7 случаев зависти, при этом все кроме одного случая оцениваются величиной менее 10%.

Таблица 5.3 – Таблица зависти

	Usa	Russia	Canada	Greenland	Norway	Iceland
Usa	0.0	5.291	-8.101	-3.166	17.809	21.513
Russia	10.992	0.0	-0.15	-0.409	7.582	15.324
Canada	10.162	3.708	0.0	-3.574	6.581	16.467
Greenland	14.966	2.332	6.989	0.0	-1.302	10.361
Norway	17.339	0.992	6.7	-0.482	0.0	8.803
Iceland	19.374	4.2	14.818	6.538	-11.574	-0.0

По этой таблице можно заметить, что у Норвегии и Гренландии свойство свободы от зависти нарушается в обе стороны. Однако ввиду того, что Норвегия владеет территориями, доступа к которым у Гренландии нет, совершение обмена между всеми территориями невозможно.

### **5.3 Анализ уязвимости социально-экономического развития Арктики из-за изменения климата**

Еще одним направлением исследований, отраженным в этом разделе отчета, является анализ уязвимости социально-экономического развития Арктики из-за изменения климата. Арктика в первую очередь испытывает на себе последствия глобального изменения климата. Изменение климата предоставляет более благоприятные условия для освоения Арктических ресурсов, что в свою очередь повышает привлекательность региона для экономической активности. Анализ этих процессов приведен в отчете.

Арктика является стратегическим регионом для многих стран, однако, некоторые из них заинтересованы в климатической и военной безопасности, а другие, как Россия, больше заинтересованы в Арктике непосредственно ради экономического благополучия страны из-за добычи и экспорта ресурсов. Вместе с этим, Арктика в первую очередь испытывает на себе последствия глобального изменения климата, ведь изменение экосистем на Севере происходит

более чем в два раза быстрее по сравнению со средними показателями по планете из-за значительных колебаний температуры, что наиболее заметно в зимний период. Задачей настоящего исследования является рассмотрение последствий изменения климата для Арктики и ее населения. Более того, важно понять, какую роль на данный момент играет экономика, основанная именно на Арктических ресурсах, и какие перспективы для экономики открываются при изменении климата. Проблема заключается в том, что изменение климата предоставляет более благоприятные условия для освоения Арктических ресурсов, что в свою очередь повышает привлекательность региона для экономической активности. Получается замкнутый круг, ведь если антропогенное влияние ускоряет таяние льдов и предоставляет возможности для освоения Арктики, то это еще больше влияет на климат.

В ходе исследования проанализированы физические и биологические процессы в Арктике, которые происходят вследствие изменения климата. По сравнению с природными процессами, доля энергии, подвластная человечеству, поистине невелика, однако перераспределение потоков уже существующей энергии может существенно влиять на климат и глобальный эффект роста температуры возникает именно из-за этого [258]. Изменения в температуре приводят к последствиям, включающим в себя изменение ледяного покрова Северного Ледовитого океана, деградацию вечной мерзлоты, повышенный риск наводнений и ускоренную эрозию берегов вследствие повышения уровня моря и усиления активности волн открытого океана, а также замещение площадей тундры другими типами растительности, включая инвазию растений [259].

Следующая часть работы посвящена влиянию изменения климата на социально-экономическое развитие населения. Больше всего подвержены этому малочисленные коренные народы, исторически проживающие в экстремальных условиях, и первыми ощущающие угрозу от эрозии берегов и наводнений вследствие того, что лед, защищающий берег, образуется позже [260]. С одной

стороны, это подталкивает жителей к климатической миграции, а с другой, перемещение пока в большинстве случаев не происходит из-за протеста самого коренного населения, привыкшего к своей среде обитаний, а также недостатка материальных средств для переезда. В любом случае, таяние льда создаст трудности для традиционного образа жизни коренного населения, основанного на рыбалке, охоте и сборе урожая [261].

Климат менялся на протяжении истории, однако ранее Арктическое население не было так зависимо от инфраструктуры, что позволяло передвигаться в более благоприятные на тот момент времени места. Сейчас же эрозия береговой линии на Аляске уже привела к тому, что дома на побережье разрушаются или оказываются в воде. Примером служит город Шишмарев с разрушенными домами из-за смещения береговой линии или город Селавик, где наблюдаются наводнения из-за близлежащих озера и реки [262]. Возникает вопрос о вынужденной климатической миграции населения, ведь потенциально под угрозой могут находиться целые Арктические деревни и города.

Наконец, в работе рассмотрено международное сотрудничество в сфере регулирования климата, а также экономическая составляющая изменений. В качестве основополагающих документов сотрудничества были выбраны Монреальский протокол, Киотский протокол, а также Парижское соглашение. Каждый из документов был проанализирован на предмет мотивации стран-участников, и тех, кто отказался от участия, а также на предмет уступок для стран, являющихся лидерами по выбросам парниковых газов. Например, при вхождении в Киотский протокол, Россия настояла на сохранении количества выбросов на уровне 1900, а не на их снижении, ссылаясь на кризис, и таким образом получила возможность торговли квотами из-за перевыполнения обязательств [263]

Практическая часть работы фокусируется на анализе экономической активности северных стран. Для анализа выбраны 8 стран, входящие в Арктический совет: Норвегия, Дания, Канада, Финляндия, Исландия, Российская Федерация, Швеция, Соединенные Штаты Америки. Основное внимание уделено

именно экономике Арктических территорий, так как для анализа важно понять, насколько страна зависит от Арктики, как пользуется ресурсами, какие существуют ограничения и возможности.

## **5.4 Оценка водообеспеченности стран Африки**

### **5.4.1 Общий анализ проблемы водообеспеченности стран Африки**

Вода, являясь важнейшим источником существования и жизнедеятельности людей, выступает фактором, провоцирующим развитие серьезных разногласий как на локальном, так и на региональных уровнях. Наличие достаточного количества водных ресурсов непосредственно определяет политическую стабильность и социально-экономическое развитие стран.

Тем не менее, само распределение водных ресурсов носит неравномерный характер. Критичное положение наблюдается в Африке, где засушливые климатические условия обострили проблему доступа населения к воде. Быстрый рост населения и как следствие увеличение водопотребления настолько повысили спрос на водные ресурсы, что потребность в воде намного превышает уровень водообеспеченности ряда африканских стран. Дефицит водных ресурсов является причиной крайней нищеты, распространения эпидемий и замедления развития стран. В то время, как изменения климата и глобальное потепление выступают катализаторами нехватки воды, отсутствие грамотного управления водными ресурсами также представляет собой серьезный вызов. Основная сложность заключается в том, что около 85% всех водных ресурсов Африки находятся в совместном управлении у нескольких стран [264]. Таким образом, поспешные и нерациональные действия одной страны могут привести к таким непоправимым последствиям для другой страны, как обострение водного дефицита, резкое повышение уровня смертности, увеличение миграционных потоков и возможное появление военных столкновений.

Актуальность данной исследовательской темы обуславливается тем, что в современном мире вода стала одним из мощнейших инструментов политических спекуляций и одним из ключевых факторов национальных конфронта-

ций. Новизна же работы состоит в рассмотрении и комплексном анализе водообеспеченности государств Африки. Несмотря на безусловную остроту поднимаемой проблемы, в зарубежной и отечественной литературе отсутствуют актуальные работы, которые бы рассматривали вопрос водообеспеченности не только в рамках одного определённого региона, но и в разрезе всего африканского континента.

В ходе работы была поставлена задача изучить явление водного кризиса и его причины в африканском регионе. Данная задача предполагала рассмотрение и анализ географических, климатических особенностей стран Африки, их водообеспеченность и водопотребление внутри стран.

#### 5.4.2 Анализ данных по водообеспечению стран Африки

Прежде чем приступать к непосредственному описанию результатов работы необходимо ввести термины и их определения, необходимые для более полного понимания проблемы:

Дефицит водных ресурсов – разница между предложением водных ресурсов и спросом на эти ресурсы. Более того, нехватка водных ресурсов может наблюдаться в водообеспеченных странах, высокий уровень водообеспеченности не исключает наличие водного кризиса [265]. В таком случае принято говорить об экономическом водном дефиците;

Улучшенный источник воды – ресурсы питьевой воды, свободные от фекального загрязнения. К данной категории относятся домашние водопроводы, общественные водоразборные колонки, буровые скважины, защищенные колодцы/родники и сбор дождевой воды [266]. Важно уточнить, что улучшенные источники воды не гарантируют, что вода будет безопасна для использования. Улучшенный лишь означает, что данный вид источника обеспечивает более безопасную питьевую воду, чем неулучшенный источник питья [267].

Если же говорить о таком понятии, как доступ населения к улучшенным источникам воды, то следует отметить, что говорить о доступе к водным ресурсам можно только в том случае, если человек живёт в пределах 1 км от улучшенного источника воды и имеет возможность получать или добывать не

менее 20 литров в день [268]. Этого объема хватит на удовлетворение базовых потребностей в питье, человеческой и пищевой гигиене.

Для определения водообеспеченности стран обычно используется индикатор водного стресса Фалькенмарка. Показатель основывается на объеме возобновляемых водных ресурсов на душу населения. Страна испытывает хронический дефицит воды, если годовой запас воды падает ниже 1000 м<sup>3</sup> на человека, если же показатель опускается ниже уровня 500 м<sup>3</sup> на человека в год, речь уже идёт о полном дефиците воды [269].

В данном подразделе отчета приведены результаты анализа данных о влиянии климатических условиях стран Африки на водообеспеченность и водопотребление, дефиците воды и его причинах. Кроме того, работа доказывает, что высокий уровень водообеспеченности не исключает наличие водного кризиса.

В ходе работы были рассмотрены 54 африканские страны и два французских заморских департамента (Реюньон и Майотта). Страны были классифицированы в соответствии с принадлежностью к водным бассейнам.

В Таблице Б.2 Приложения Б приведено разбиение стран Африки на группы по степени удовлетворенности их водными ресурсами в соответствии с индикатором водного стресса Фалькенмарка.

В качестве основных причин абсолютного дефицита водных ресурсов, наблюдающегося в четырех африканских странах, выступают климатические условия и особенности гидрографической сети: в данных странах отсутствуют постоянные водотоки, большая часть водоёмов сезонные, представленная уездами. Однако дефицит водных ресурсов, с которым столкнулся Алжир, достаточно уникален по сравнению с другими странами Африки. Нехватка воды вызвана не столько климатическими изменениями, сколько быстрым процессом урбанизации. Около 74% населения проживают в городах [270]. Массовая миграция из сельских районов в города выступила в роли катализатора усиливающегося давления на городскую водоснабжающую и канализационную инфраструктуру. Так, жители столицы имеют доступ к водным ресурсам всего

30% времени [271]. Не менее значительным препятствием на пути обеспечения населения безопасной питьевой водой в Алжире является неэффективная и устаревшая инфраструктура. Разрывы в трубах, постоянные утечки являются причиной, из-за которой до 30% водных ресурсов теряются в пути [272]. Ситуация с водными ресурсами в Тунисе и Джибути осложняется благодаря низкому качеству воды ввиду высокой концентрации соли. Так, подземные воды хорошего качества можно найти только в 8% мелких и 20% глубоких водоносных горизонтах. Около же 36% ресурсов подземных вод не пригодны для использования в сельскохозяйственном секторе и в производстве безопасной питьевой воды, поскольку содержание соли превышает допустимый уровень в 3 г/л [273].

Ряд стран, хорошо обеспеченных водными ресурсами, к примеру, Демократическая республика Конго и Ангола, испытывают так называемый экономический водный кризис из-за ограниченного доступа людей к воде. Низкий уровень доступа населения к воде, в свою очередь, может определяться недостатком очистных сооружений или полным отсутствием канализационных систем, что провоцирует распространение заболеваний, передающихся через воду, и, как следствие, низкое качество воды. Так, в трёх основных реках Анголы – Бенго, Данде и Кванза, была зафиксирована высокая концентрация кишечной палочки [274]. Кроме того, в качестве ещё одной причины ограниченного доступа к водным ресурсам выступает большое расстояние до источников безопасной питьевой воды. В свою очередь, расстояние до источников воды, а именно близость или само наличие воды из водопроводной системы, оказывает влияние на ценообразование в сфере обеспечения населения водой: цены растут по мере увеличения расстояния до источника воды и по мере усложнения дорожных условий в сезон дождей. Существует колоссальная разница в ценах в зависимости от использования того или иного вида источника. Так, домохозяйства, покупающие воду у местных поставщиков в пригородных районах, платят в 10 раз больше, чем те домохозяйства, которые имеют непосредственное прямое водопроводное подключение к дому. Ввиду того, что

прямое подключение менее распространено в районах, расположенных южнее Сахары, и свойственно для богатых домохозяйств, менее богатые домохозяйства платят значительную часть своего бюджета на водопотребление: наиболее бедным домохозяйствам приходится тратить около 15-20% доходов на покупку воды. Такая дороговизна водных ресурсов приводит к изменению потребительского поведения, неизменно сокращая потребление воды семьями или провоцируя их перенаправлять свои средства на удовлетворение таких базовых потребностей, как медицина и питание [275].

Не менее серьезной причиной дефицита водных ресурсов является рост населения и следующее из этого увеличение давления на окружающую среду, что приводит к переиспользованию водоносных горизонтов и к их последовательному загрязнению. Так, в результате стремительного процесса увеличения численности населения около 600 миллионов кубических метров хозяйственно-бытовых сточных вод ежегодно производятся и выбрасываются в окружающую среду прибрежных районов Марокко необработанными [276].

Глобальное потепление и изменения климата также оказывают значительное влияние на водообеспеченность стран и на наличие доступа к безопасной питьевой воде. Так, одним из главных препятствий на пути решения водного кризиса в Южноафриканской республике является увеличение объемов осадков, а именно увеличение интенсивности дождей, что является причиной вымывания отходов в водотоки и высокого загрязнения воды, которое представлено большой концентрацией кишечной палочки [277].



## **5.5 Продовольственная безопасность ряда африканских стран и её зависимость от изменения климата**

### **5.5.1 Общий анализ проблемы продовольственной безопасности африканских стран**

На сегодняшний день Африка является одним из наиболее уязвимых континентов перед изменением климата. Последствия изменения климата затрагивают все сферы жизнедеятельности населения: здравоохранение, продовольственную безопасность, процессы миграции, энергетический комплекс и различные социальные аспекты. Большое количество публикаций посвящено изучению данного вопроса [278]- [280], что подчёркивает актуальность исследования влияния изменения климата на экономическое и социальное положение стран африканского континента.

По причине своего географического расположения Африку пересекают несколько климатических поясов, а именно экваториальный, субэкваториальный, тропический и субтропические пояса. Такое климатическое разнообразие привлекает больший интерес в изучении проблем продовольственной безопасности, поскольку для одних стран последствия от изменения климата могут оказаться отрицательными, для других они могут нести положительный характер. Аграрный сектор является одним из самых главных секторов африканской экономики, на 2019 год аграрный сектор обеспечивает около 60% рабочих мест для африканцев [281] и является одним из наиболее чувствительных секторов экономики к изменению климата. Состояние и производительность аграрного сектора влияют не только на экономическое положение континента, но и на благосостояние населения в целом, в частности на продовольственную безопасность.

Прежде чем перейти к рассмотрению влияния изменения климата на продовольственную безопасность, следует изучить текущие климатические условия на территории Африки. Основными индикаторами изменения климата являются колебания средних температур и количества осадков, по-

сколькx эти два показателя в большей степени оказывают влияние на производительность аграрного сектора. По состоянию на 2019 год зарегистрирован рост средних температур на 2°С по сравнению с уровнем средних температур за период 1981-2010 гг. на территориях Южно-Африканской Республики, Намибии и в некоторых частях Анголы [281]. На остальных территориях наблюдается повышение средних температур на 1° С по сравнению с 1981-2010 гг. [281]. По прогнозам Межправительственной группы экспертов по изменению климата на африканском континенте в последующие 100 лет ожидается увеличение температур на 2-6 ° С, а также увеличение изменчивости осадков [282]. К 2050 году прогнозируется, что средние температуры в Африке увеличатся на 1.5-3 ° С [283].

Цель исследования заключается в проведении анализа влияния изменения климата на продовольственную безопасность ряда африканских стран с 1961 по 2016 гг.

#### 5.5.2 Аналитический обзор научной литературы по исследованию продовольственной безопасности стран Африки

Как было отмечено ранее, аграрный сектор экономики также подвержен отрицательным климатическим изменениям. Этот сектор является одним из важнейших, так как производство сельскохозяйственной продукции является одним из главных источников дохода у африканских семей. Климатические изменения, а именно колебания температур и количества осадков, негативно сказываются на качестве и количестве урожая, тем самым ставя под угрозу благосостояние населения. Ввиду сокращения урожайности из-за изменения климата, возрастают цены на сельскохозяйственную продукцию, что усугубляет проблему недоедания и повсеместного голода.

Многие ученые [284, 285, 286] посвятили ряд работ выявлению взаимосвязи между урожайностью и климатическими изменениями. В [285] для изучения ранее указанной взаимосвязи для африканских стран к югу от пустыни Сахары в период с 1961 по 2002 гг. авторы используют регрессионный анализ.

В качестве регрессоров используются параметры, характеризующие ежемесячные данные по изменению температур и количеству выпавших осадков за указанный промежуток времени и страновой фиксированный эффект. В качестве зависимой переменной используется урожайность для каждой африканской страны к югу от пустыни Сахары (тонн/га) пяти основных культур: кукурузы, сорго, просо, арахиса и маниока. Выбор данных культур объясняется, тем что эти культуры содержат высокий уровень белка и жира, что важно для правильного рациона питания человека. Авторы пришли к выводу, что действительно существует отрицательная зависимость между климатическими изменениями и урожайностью. Наибольшее негативное влияние на урожайность оказывает повышение температур, однако не на все культуры. На урожайность маниоки изменение погодных условий несильно сказывается, поскольку этот корнеплод обладает сильно изменяющимся вегетационным периодом. Таким образом, увеличение температур может пагубно отразиться на производительности аграрного сектора экономики африканских стран.

В [284] отмечается, что сокращение урожайности кукурузы и сорго в большинстве стран восточной Африки соответствует уменьшению количества осадков на 50-150 мм в период 1996-2003 гг. При увеличении глобальных температур на  $4^{\circ}\text{C}$  прогнозируется сокращение урожайности на 24% для кукурузы и 71% для бобов в странах к югу от Сахары [286].

Увеличение температур также вызывает ряд проблем, связанных с качеством потребляемых африканцами продуктов питания, а именно существует мнение, что с ростом температур, продовольственные продукты начинают быстрее портиться, что отрицательно сказывается на процессах пищеварения, обостряя проблему повсеместного голода и недоедания. Например, при увеличении средних температур на  $1.2-1.7^{\circ}\text{C}$  к 2050 году количество недоедающего населения к югу от пустыни Сахара увеличится на 25% процентов [287]. Недоедание, в свою очередь, может неблагоприятно отразиться на росте и развитии

детей. Если к 2050 году будет зарегистрировано увеличение средних температур на 1.2-1.7° C, то данное изменение приведёт к увеличению количества детей с отставанием в росте на 9%.

## **5.6 Применение интервального оболочечного анализа данных к оцениванию эффективности превентивных противопожарных мероприятий**

### **5.6.1 Постановка задачи**

К сожалению, чрезвычайные ситуации, как природные, так и техногенные, являются неотъемлемой частью современного мира. Они постоянно сопровождают людей, угрожают их жизни, приносят боль и страдания, наносят ущерб и разрушают материальные ценности, причиняют огромный, зачастую непоправимый ущерб окружающей среде, обществу и цивилизации.

Ежегодный рост числа жертв на 8-9% и экономических потерь на 10% в результате чрезвычайных ситуаций является устойчивой тенденцией [288]. Глобальный ущерб от стихийных бедствий может составить около 160 миллиардов долларов ежегодно. Кроме того, масштабы и сложность технологических процессов стабильно растут, используются радиоактивные и химически опасные вещества. Все эти факторы подчеркивают важность проблем, связанных с обеспечением безопасности, сохранением экономического потенциала и окружающей среды в чрезвычайных ситуациях.

Поскольку важно прогнозировать и смягчать последствия катастроф, возникает вопрос, как это правильно выполнить в определенных условиях. Учитывая, что до сих пор не существует единых правил, единственным решением представляется простое повторение наиболее удачных примеров. По этой причине крайне важно определить, какие примеры являются эффективными, а какие нет.

Поэтому, необходимо применять некоторые методы оценки эффективности, сравнивать результаты на разных примерах и выбирать наилучшую альтернативу в качестве эталона.

Поскольку в [289] утверждают, что количественная оценка очень чувствительна к важности различных факторов, было решено использовать подход линейного программирования для оценки эффективности. Такой подход был предложен в [290] и состоит в использовании линейной функцией доли несколько особенностей объекта, так как ее эффективность. Позднее этот подход превратился в широко используемую методологию, известную как оболочечный анализ данных [291]. В настоящее время он используется в разных сферах и для разных задач: эффективности формирования финансового портфеля [292], контроля парниковых газов в производстве электроэнергии [293], сравнения производственных фирм [294] и многих других.

Оболочечный анализ данных (DEA) основан на идее оценки эффективности различных объектов принятия решений (decision making unit – DMU) с помощью отношения входных и выходных параметров объекта. Все объекты характеризуются двумя векторами – «входами», такими как затраченные ресурсы, и «выходами», которые обычно интерпретируются как достигнутые результаты.

Рассматривая многоцелевую оптимизационную задачу (минимизацию затрат и максимизацию выходных результатов), [291] предложили рассчитать эффективность DMU как отношение взвешенной суммы «выходов» к взвешенной сумме «входов». Кроме того, с учетом рациональности и интерпретируемости эффективности необходимы ограничения, гарантирующие, что эффективность всех объектов лежит в интервале  $[0,1]$ .

В результате мы получаем следующую оптимизационную задачу:

$$\max_{u_i, v_j} \frac{\sum_{i=1}^M u_i y_{ik}}{\sum_{j=1}^N v_j x_{jk}}.$$

С учётом ограничений  $\forall i, j, k$

$$\begin{cases} \frac{\sum_{i=1}^M u_i y_{ik}}{\sum_{j=1}^N v_j x_{jk}} \leq 1, \\ u_i \geq 0, \\ v_j \geq 0, \end{cases}$$

где  $y_{ik}$  и  $x_{jk}$  – это выходные и входные параметры  $k$ -го DMU, в то время как  $u_i, v_j$  – неотрицательные коэффициенты, показывающие важность каждого показателя в отдельности.

Решая эту задачу для каждого объекта из выборки, мы получаем оптимальную границу, где эффективность равна 1. Для всех DMU, лежащих ниже этой границы, эффективность оценивается с учётом расстояния от этой эталонной границы.

DEA применяется для анализа различных катастроф. Например, в [295] применили DEA к оценке уязвимости от наводнений в районе озера Дунтин (Китай). В [296] проанализировали техногенные катастрофы в различных регионах Российской Федерации. Между тем в [297] предложили использовать его для анализа эффективности реализации политики снижения риска землетрясений, а в [298] работали с уязвимостью важных экономических регионов Китая к тайфунам.

Кроме того, существуют некоторые модифицированные методы оболочечного анализа данных. Например, в [299] использовали DEA для оценки уровня безопасности на атомной электростанции. Для этой цели было предложено учесть спецификацию некоторых проблем. Так, было заявлено, что вероятность режима отказа является гораздо более важным показателем, чем все остальные критерии. Поэтому необходимо наложить некоторые ограничения на весовые коэффициенты признаков. Это позволит построить более реалистичный и более точный метод, который будет обращать внимание на соотношение важности различных критериев, которые необходимо соблюдать при решении тех или иных задач.

Однако для применения всех вышеупомянутых версий DEA необходимо получить точные оценки всех функций DMUs. Между тем некоторые характеристики, такие как население региона или ВВП, оцениваются примерно, поскольку небольшие отклонения в этих параметрах не столь важны.

Кроме того, есть некоторые характеристики, которые не могут быть измерены непосредственно. Например, в случае сравнения эффективности мер

по предупреждению стихийных бедствий в регионе с помощью некоторых имитационных моделей оцениваются такие характеристики, как потенциальное число погибших или раненых, или общие экономические потери. Поэтому эти параметры не могут быть точными из-за неточности процесса моделирования и обычно даются в виде приближенных значений.

В результате становится ясно, что для применения в реальной жизни лучше использовать некоторые специфические модификации DEA, которые могут работать с приближительными данными.

### 5.6.2 Основная модель

В рамках исследования была предложено два новых метода, позволяющих решить проблему неточности в данных. Так, мы предлагаем использовать для оценки параметров простые интервалы вместо точечного значения (пара  $(y_{ik}^-, y_{ik}^+)$  вместо  $y_{ik}$ ). И поскольку в оболочечном анализе данных необходимо сравнивать объекты, мы вводим новую методологию сравнения значений параметров ( $>_i$  – сравнение по  $i$ -му выходному признаку):

$$\text{объект}_k >_i \text{объект}_l \Leftrightarrow y_{ik}^- > y_{il}^+.$$

В свою очередь, если выполняются неравенства  $y_{ik}^+ > y_{il}^-$  и  $y_{il}^+ > y_{ik}^-$ , то это означает, что интервалы  $(y_{ik}^-, y_{ik}^+)$  и  $(y_{il}^-, y_{il}^+)$  пересекаются, и объекты  $k$  и  $l$  несравнимы.

Используя этот тип представления данных и сравнение параметров, мы предлагаем использовать два метода IDEA (интервальный DEA).

Первый основан на идее, что некоторые DMU могут находиться вблизи границы эффективности. Но, в случае классического DEA, они не получают 100% эффективности. Мы хотим избавиться от этого недостатка. Следовательно, мы предлагаем присвоить 100%-ую эффективность не только объектам на границе наилучшей эффективности, но и тем, которые с ними несравнимы. В результате строится так называемая "идеальная трубка" из 100% эффективных объектов (следовательно, метод называется идеей "идеальной

трубки"). Эффективность всех остальных DMU оценивается классическим DEA.

Второй предложенный метод основан на идее того, что любой параметр (как входной, так и выходной) может быть наиболее важным при сравнении объектов. Следовательно, если один из объектов имеет наилучшую ценность хотя бы по одному признаку, то он должен рассматриваться как наилучший. Технически это означает, что следует использовать принцип оптимальности Парето:

$$\text{объект}_k > \text{объект}_l \Leftrightarrow \begin{cases} y_k \geq y_l \\ x_k \leq x_l \end{cases} \Leftrightarrow \begin{cases} \forall i y_{ik}^- \geq y_{il}^+, \\ \forall j x_{jk}^+ \leq x_{jl}^-. \end{cases}$$

Этот принцип обладает некоторыми привлекательными свойствами [300] и может быть эффективно использован для сравнения объектов с интервальными оценками признаков.

Согласно принципу оптимальности Парето, множество наилучших объектов строится из всех объектов, которые не являются доминируемыми по Парето

$$C_{Pareto}(X) = \{y \in X | \nexists x \in X: x > y\}.$$

В результате мы получаем Парето-модификацию оболочечного анализа данных, согласно которой на первом этапе процедура выбирает все Парето-оптимальные объекты и присваивает им 100%-ую эффективность. После этого все остальные объекты оцениваются классическим DEA.

После разработки данных методов мы сравнили профилактические меры против лесных пожаров в разных регионах Российской Федерации.

Безусловно, для организации любых мер предосторожности требуются финансовые средства. Однако информация о прямых расходах на противопожарные мероприятия отсутствует, поэтому в качестве входных параметров мы использовали два разных денежных потока: на охрану окружающей среды и инвестиции в сельское и лесное хозяйство.



Кроме того, необходимо также параметризовать результаты профилактических мер. Для этого мы использовали два основных показателя: количество природных пожаров в регионе и площадь лесных угодий, охваченных пожарами.

Для того чтобы провести наше исследование более корректно, мы сравнивали только 46 российских регионов, которые действительно сталкиваются с проблемой природных пожаров (в этих регионах происходит не менее 10 природных пожаров в год).

В результате применения различных методов оболочечного анализа данных были получены три оценки эффективности для каждого региона (классический DEA и 2 модифицированных метода). Другими словами, получилось три различных рейтинга российских регионов, в соответствии с каждой модификацией DEA.

Как и ожидалось, результаты всех методов дают нам близкие упорядочивания с небольшими локальными различиями. Переходя к конкретным результатам, необходимо отметить семь регионов, которые получают 100%-ую эффективность согласно всем трём методикам: Новгородская область, Удмуртия, Пензенская область, Ульяновская область, Республика Хакасия, Камчатский край и Магаданская область. Хорошие результаты этих регионов можно объяснить разными причинами. Первая пятерка – это просто регионы с небольшой территорией (менее 0,4% площади Российской Федерации), поэтому контролировать и противостоять лесным пожарам несложно. Магаданская область в основном состоит из горных пустынь, поэтому лесные пожары также не являются серьезной проблемой для этого региона. В свою очередь, Камчатский край – единственный регион, который действительно можно охарактеризовать как эталон, поскольку более 14,5% территории Камчатского края представляет собой особо охраняемую территорию, и, согласно результатам проведенных исследований, охрана обеспечивается высококвалифицированными специалистами.

Что касается худших регионов, следует упомянуть Красноярский край, Иркутскую область и Свердловскую область. Однако, их низкий рейтинг можно объяснить тем, что все эти регионы имеют очень большую площадь (например, Красноярский край занимает около 14% площади Российской Федерации).

### **5.7 Некоторые выводы**

Исходя из работы по миграции в арктических регионах следует, что мигранты вносят значительный вклад как в принимающую, так и в родную страны. Миграция – это растущая глобальная реальность, с 2000 года количество мигрантов выросло с 2,8% до 3,4% населения мира. Международное сотрудничество имеет важное значение, поскольку миграция, скорее всего, будет продолжать расти. Более того, мигранты вносят значительный вклад в экономический рост.

В будущем будут присутствовать как экзогенные, так и эндогенные демографические проблемы, основанные на текущих тенденциях и будущих прогнозах. Некоторые проблемы уже решаются с помощью различной демографической политики в Арктике. В Нунавуте, Гренландии и Фарерских островах есть конкретные программы по предотвращению роста урбанизации.

Демографическая проблема старения возникает во всем мире, и Арктический регион не исключение. Высокий уровень оттока молодого населения – это тенденция многих небольших арктических поселений. Миграция в Арктику в значительной степени зависит от международной миграции, которая препятствует старению и сокращению населения.

Изменение климата и таяние вечной мерзлоты – экзогенные факторы, влияющие на население Крайнего Севера. Морской лед сокращается, и это вызывает необходимость переселения прибрежных поселений на Аляске. Таяние вечной мерзлоты сказывается на инфраструктуре населенных пунктов по всей Арктике. Уменьшение количества льда в океане в Арктике открывает новые

возможности для судоходства и добычи полезных ископаемых и имеет как положительные, так и отрицательные последствия для портовых арктических городов.

Основные факторы, касающиеся Арктики, неопределенны: спрос и цена на ресурсы Арктики, влияние глобального потепления на регион в целом, инновации и принятие новых технологий обществом, будущее сообществ коренных народов. Вот почему становится почти невозможным определить демографические направления, занятость в Арктике, судьбу деревень и наиболее эффективную административную политику для Крайнего Севера.

Вспышка коронавируса показала нам, что нас ждет неопределенное будущее. Изменения, с которыми сталкивается Арктика, являются одними из самых серьезных в мире. Для решения социально-экономических, климатических и миграционных проблем арктического региона потребуются активные политические действия и эффективная разработка политики.

Результаты данного исследования могут применяться для принятия решений по экономической и демографической деятельности в местных органах управления Арктики.

В рамках работы по распределению спорных территорий арктического региона были рассмотрены существующие методы решения задачи территориального распределения и предложены новые. Была разработана программная реализация методов, применимая к любой другой задаче территориального дележа, учитывающая различный уровень заинтересованности участников дележа в ресурсах и связность территорий. Практическое применение модели на реальных данных арктического региона дало результат итогового распределения свободных участков, в котором уровни удовлетворённости всех стран оказались практически равны, а территории связаны. Данная модель может помочь с решением других, в том числе реальных территориальных задач, ведь своевременный анализ проблемы с учётом различных моделей, подходов и сценариев облегчает процесс принятия таких трудных и важных решений.

Результаты данной работы нашли отражение в дипломной работе «Распределение спорных территорий с применением математических моделей на примере Арктического региона».

Рассмотрены механизмы и масштаб изменения климата в Арктике, в первую очередь через антропогенное воздействие, из-за которого и ускорились темпы потепления. Проведенный анализ, в частности, показывает, что для подписания климатических соглашений, регулирующих парниковые выбросы, необходимо идти на существенные уступки. На данном этапе работы осуществляется сбор данных по экономике Арктических стран, необходимый для верификации модели.

В соответствии с целью работы по оценке водообеспеченности стран Африки – анализ причин водного дефицита, были изучены 56 стран Африки, что позволило понять основные факторы, выступающие в качестве препятствий на пути обеспечения населения безопасной питьевой водой.

Основными причинами водных кризисов являются изменение климата, увеличение численности населения, и как следствие увеличение водопотребления, которое превышает уровень водообеспеченности в ряде африканских стран. Кроме того, анализ стран позволил понять, является ли большой запас водных ресурсов гарантом отсутствия водного дефицита. Страны, хорошо обеспеченные водными ресурсами, к примеру, Демократическая республика Конго и Ангола, столкнулись с серьезнейшим водным кризисом из-за низких показателей доступа населения к безопасной питьевой воде. Ограниченный доступ населения к водным ресурсам, как уже упоминалось ранее, может быть объяснен бедными и устаревшими водоснабжающей и канализационной инфраструктурами, что, в свою очередь, ведет к распространению заболеваний, передающихся через воду. Не менее важным препятствием по обеспечению населения водой также является повышение уровня моря, что провоцирует вторжение соленой воды в пресные, пригодные для питья, водотоки.

Оценивая полноту решения поставленных задач, необходимо отметить, что задачи были осуществлены в полной мере. Тем не менее, в ходе анализа

водообеспеченности, водопотребления и доступа населения к водным ресурсам, мы столкнулись с рядом ограничений. Так, отсутствие современной и актуальной информации или ее полное отсутствие в ряде случаев осложняет процесс предоставления точной и комплексной картины происходящего.

Результаты работы нашли отражение в курсовой работе «Modeling of Tensions Among Countries in North Africa on the Bases of Water Resources Deficit» и в докладе «Оценка водообеспеченности стран Африки».

Данная работа может стать отправной точкой для более конкретного и детального рассмотрения проблемы водных ресурсов и стать основательной базой для изучения влияния наличия водных ресурсов на миграционные потоки и на политико-экономические отношения стран-соседей.

Как итог исследовательской работы касательно продовольственной безопасности Африки можно заметить, что Африка является наиболее чувствительным континентом к климатическим изменениям, в частности к изменениям температур и количеству осадков. На сегодняшний день средние годовые температуры имеют тенденцию расти, что отрицательно сказывается на производительности аграрного сектора экономики. Данные последствия напрямую отражаются на качестве и количестве продуктов питания, уровне недоедания и повсеместного голода. По этой причине исследования в области продовольственной безопасности являются наиболее важными.

На данном этапе исследование представляет собой анализ текущей климатической и продовольственной ситуации на африканском континенте. В дальнейшем будут проанализировано влияние изменения температур и количества осадков на производительность аграрного сектора ряда африканских стран с 1961 по 2016 гг.

В рамках разработки модификаций оболочечного анализа данных были разработаны новые модификации этого метода, позволяющие работать с неточными данными.

Новые методы были апробированы на решении задачи оценки эффективности превентивных противопожарных мероприятий в регионах Российской Федерации [301]. Обе разработанные модификации решают проблему не точности данных с помощью интервальных оценок. Полученные результаты оказались легко интерпретируемы и не сильно отличаются от решений классического оболочечного анализа, что показывает практическую эффективность новых методов.

## **6 Исследование моделей регулирования, распространения и специализации**

В этот раздел отчета вошли результаты исследований некоторых моделей регулирования, распространения и специализации, а именно:

- исследование моделей регулирования в банковской сфере: подход на основе внутренних рейтингов для расчета требований к капиталу;
- исследование моделей специализации в абстрактных системах: случай структурных ограничений, задаваемых неравенствами;
- разработка матричного метода линеаризации систем уравнений в частных производных на постоянных решениях.

### **6.1 Исследование моделей регулирования в банковской сфере: подход на основе внутренних рейтингов для расчета требований к капиталу**

#### **6.1.1 Постановка задачи**

В 2013 году объем теневого сектора в мире занял 120% от мирового ВВП [302]. В абсолютном выражении это составило 75 трлн \$, где 2/3 от этого пришлось на США и Европейский Союз. Размер теневого сектора в России на порядок меньше – 0,08 трлн \$. По данным [303] после 2013 года объем теневого сектора продолжает расти. Теневой банковский сектор характеризуется агрессивным поведением банков с точки зрения риска наряду с игнорированием требований регулятора и препятствию надзора. Это увеличивает вероятность кризиса и потери при нем, что напрямую влияет на уровень благосостояния общества. В доказательство этому есть несколько примеров из практики, когда мошенническая деятельность банков вела к существенным потерям. Например, Citi, Goldman Sachs, BCCI, Barrings Bank, Johnson Matthey Bank, банк «ФК Открытие» [304], Societe Generale.

Несмотря на это, модель [305], опубликованная в American Economic Journal, показывает иной вывод: скрывание риска в рамках теневого банкинга повышает благосостояние общества, и его наличие в экономике оправдано. Работа строится на спорных предпосылках, что может влиять на устойчивость

этого вывода. Поэтому цель настоящего исследования – обсудить предпосылки модели [305] и проверить, как изменяется вывод о влиянии теневого банкинга на благосостояния общества после корректировки допущений.

В настоящей работе ключевым изменением в модели [305] предлагается добавление ограничений в виде требований к капиталу. Рассчитаны границы нормативов достаточности капитала (CAR), превышение которых может увеличивать вероятность перехода банков в теневой банкинг и их внебалансовое кредитование. Такой вывод показывает в частности, что регулирование, предложенное в THE MINNEAPOLIS PLAN<sup>4</sup> и предполагающее повышение требований к CAR до 23.5%, и другие инициативы по значительному увеличению требований, чтобы снизить влияние крупных банков, может выполнять обратную поставленную задачу – наращивать риск внутри теневого (невидимого) регулятору банкинга.

#### 6.1.2 Описание анализируемой модели

В модели [305] существуют:

- три типа активов: безрисковые, низкорискованные и высокорискованные;
- три агента: инвесторы, регулятор и банки<sup>5</sup>.

Модель содержит ряд спорных предпосылок. Во-первых, в работе не введено ограничение на капитал, которое влияет на выбор портфель активов банками и выбор типа банкинга, на котором осуществлять кредитование. С одной стороны, высокие требования сподвигнут больше банков переходить на теневой банкинг. С другой стороны, за счет требований к капиталу размер потерь в традиционном (нетеневом) банкинге меньше. Оба момента влияют на итоговый результат, каких объемов достигнет теневой банкинг. Вопрос связи

---

<sup>4</sup> <https://www.minneapolisfed.org/~media/files/publications/studies/endingtbtf/the-minneapolis-plan/the-minneapolis-plan-to-end-too-big-to-fail-2016.pdf?la=en> (план упомянут в [306]).

<sup>5</sup> В [305] четко не указано, чем являются средства инвесторов – вкладами или долями участия в капитале.



требований капитала и размеров теневой банковской системы является актуальным. Исследователи в целом не пришли к однозначному выводу о том, требуется повышать или снижать требования к капиталу, чтобы повысить финансовую стабильность [307]. Однако можно видеть, что политика США склонна к повышению требований. Так, например, в 2016 году опубликован THE MINNEAPOLIS PLAN<sup>6</sup> (упомянуто в [306]). План предполагает снизить влияние системно значимых банков (в том числе теневых банков, которые оказывают высокое влияние на экономику) за счет повышения требований к CAR до 23,5%. Настоящая работа поможет привести дополнительные аргументы в пользу мнения, что требования к капиталу не могут быть завышены, так как при достижении определенной границы CAR это может привести к увеличению теневой системы.

Во-вторых, [305] приуменьшает роль и возможности регулятора оценивать риск банков и системы в целом. Автор считает, что регулятор может дифференцировать только безрисковые и рискованные активы, но в группе рискованных активов не умеет различать активы с низким и высоким уровнем риска. Невозможность регулятора в модели [305] предотвратить вложения в активы с высоким уровнем риска, заставляет его вводить запрет на все виды рискованных активов. Такая ограничительная политика не учитывает возможность регулятора оценить риск вложения банков через внедрение продвинутых подходов к оценке риска (подход на основе внутренних рейтингов для оценки величины кредитного риска, ПВР), хотя данная возможность активно используется регулятором, в том числе, в США, чья политика известна [305]. Так в 2014 году 15 ключевых банков США, которые составляют 75% активов системы, перешли в обязательном порядке на ПВР. Другим банкам был предоставлен выбор в использовании этой опции. В России к 2019 году на ПВР перешли 2 системно

---

<sup>6</sup> <https://www.minneapolisfed.org/~media/files/publications/studies/endingbtbf/the-minneapolis-plan/the-minneapolis-plan-to-end-too-big-to-fail-2016.pdf?la=en>

значимых банка, объем активов которых составляет примерно 34% на 01.01.2020. Более того в 2019 году Банк России объявил об обязательности перехода на ПВР всех СЗКО (78% активов системы на 01.01.2020).<sup>7</sup> Режим ПВР обязывает банки использовать продвинутые подходы к оценке риска и публиковать отчетность, чтобы раскрывать уровень принятого риска, его оценки и сравнение с фактическим реализованным риском. Это означает, что регулятор и инвесторы имеют возможность определить уровень риска, который банк на себя принимает. Учет режима ПВР может изменить выводы [305] о переходе банков в теневой банкинг.

#### 6.1.3 Включение режима ПВР и регулирования капитала в анализируемую модель

О модификациях модели [305].

В качестве первой модификации модели [305] добавлен учет регулирования капитала через норматив, который требует от банков обладать достаточным количеством капитала, чтобы быть способными покрыть непредвиденные потери.

Введение норматива достаточности капитала повлияет на выбор банка объема и риска вложений. При увеличении норматива банк будет либо понижать объем вложений ( $A_i$ ), либо переходить к менее рискованным активам (с более низким  $RW_i$ ). Поскольку ожидаемая доходность низкорисковых активов (LR, low risk) выше, чем у безрисковых активов (NR, no risk), банки будут выбирать первую стратегию – понижать объем вложений.

В качестве второй модификации модели [305] добавлен альтернативный режим расчета кредитного риска – подход на основе внутренних рейтингов (ПВР). В отличие от стандартизированного подхода, в котором банк полагается на оценки  $RW_i$  регулятора, ПВР позволяет точнее оценивать риск банку

---

<sup>7</sup> По данным kuap.ru

на основе результатов собственных внутренних моделей. Благодаря более чувствительной оценке риска, которая затем отражается в отчетности, регулятор видит достаточно дифференцированный разрез уровня риска банков.

Внедрение ПВР моделей регулятором дает возможность ему надзирать за уровнем риска активов, а следовательно, различать активы с низким и высоким уровнем риска. Это позволяет регулятору вести менее ограничительную политику. Регулятор все еще запрещает инвестировать в активы с высоким уровнем риска, но отменяет запрет на вложения в активы с низким уровнем риска в рамках традиционного сектора. Одновременно регулятор требует соблюдения норматива достаточности капитала (что не учтено в [305]).

В модели такое изменение для банков в политике регулятора, с одной стороны, расширяет круг возможностей инвестирования за счет снятия запрета на вложения в низкорисковые активы, с другой стороны, ограничивает вложения через норматив. В этом случае банк будет решать следующую задачу. Если к активам с низким уровнем риска есть доступ, то банк решает, в каком банкинге (традиционном или теневом) инвестировать в низкорискованные активы. Иначе, при отсутствии доступа, банк решает, вкладываться в безрисковые активы в традиционном (нетеневом) банкинге или в высокорискованные активы в теневом банкинге.

О формировании инвесторами ставки процента за предоставления средств.

Банки, применяющие ПВР, имеют официальное публичное разрешение регулятора, что позволяет инвесторам их идентифицировать. Таким образом, ставка для ПВР банков, кто вкладывается в низкорискованные активы, будет обратно пропорциональна доле инвестиций за вычетом ожидаемых потерь по ним. Для остальных типов активов возможно рассмотреть два варианта формирования ставок, как представлено в Таблица 6.1.

Таблица 6.1 – Методы формирования ставок инвесторов

Тип банкинга	Актив	Вариант 1	Вариант 2
TB	LR	$R_{TB-LR}$	$R_{TB-LR}$
	NR	$R_{TB-NR}$	$R_{rest}$
SB	LR	$R_{SB}$	
	HR		

6.1.3.1 Выбор между вложением в безрисковые активы в традиционном (нетеневом) банкинге или в высокорискованные активы в теневом банкинге.

В первую очередь банк выбирает между вложениями в безрисковые активы в традиционном (нетеневом) банкинге или в высокорискованные активы в теневом банкинге. Для этого он сравнивает уровень прибыли при участии в каждом из вариантов. В случае традиционного банкинга вложения банка ограничены требованием к капиталу (модификация № 1).

Разница в выручке двух случаев (первое слагаемое в формуле) в терминах [305] будет соответствовать издержкам теневого банкинга:

$$C_{SB} = (Inv + k)((1 - EL_{NR})\overline{Y}_{NR}R_{NR}^L - (1 - EL_{HR})R_{HR}^L). \quad (6.1)$$

Разница в выплатах инвестору будет представлять выгоду от теневого банкинга:

$$B_{SB} = Inv \left( (1 - EL_{NR})\overline{Y}_{NR}R_{TB-NR} - (1 - EL_{inf})R_{SB-HR} \right) - k(\overline{Y}_{NR} + 1). \quad (6.2)$$

Равенство издержек и выгоды определит оптимальный уровень участия банков в теневом банкинге при вложении в высокорискованные активы.

Если ставка инвесторами формируется по варианту 1, то  $R_{TB-NR}$  определяется ожидаемыми потерями, а  $R_{SB-HR}$  зависит от выбора банков, в каком типе банкинга и в каком объеме вкладываться в LR активы ( $\overline{x}$ ).

Поскольку выбор между инвестициями в NR или HR активы зависит от выбора объемов и типа банкинга инвестирования для LR активов, рассмотрим, как этот выбор делается.

6.1.3.2 Выбор между традиционным и теневым банкингом для вложений в низкорискованные активы.

Регулирование капитала также как и в случае с NR активами будет ограничивать количество инвестиций, который банк может сделать в низкорисковые активы в традиционном (нетеневом) банкинге.

В терминах [305] рассчитаем издержки и выгоды от теневого банкинга при выборе между вложениями в LR активы в традиционном или теневом банкинге. Под издержками теневого банкинга понимается разница между выручкой в традиционном и теневом банкинге. Ранее определили, что издержки будут равны:

$$C_{SB-LR} = (1 - EL_{LR})(Inv + k)R_{LR}^L(\alpha - \overline{\gamma_{LR}}). \quad (6.3)$$

Под выгодой теневого банкинга понимается разница между издержками банка в традиционном и теневом банкинге.

$$B_{SB-LR} = (1 - EL_{LR})Inv(\overline{\gamma_{LR}}R_{TB-LR} - \alpha R_{SB-LR}) + (\alpha - \overline{\gamma_{LR}})k. \quad (6.4)$$

Вариант 1 формирования ставок инвестора.

Ставки в традиционном и теневом банкинге будут равны только в случае, если ни один из банков не вложиться в NR активы.

На Рисунок 6.1 показано, что уровень прибыли в традиционном секторе не зависит от доли вложения в NR активы (при формировании ставки инвесторами по варианту 1), но будет снижаться при увеличении норматива CAR. Уровень прибыль в теневом секторе снижается с ростом доли в LR в традиционном (нетеневом) банкинге. В этом случае инвестор видит, что теневой сектор в большей степени заполняется NR активами и повышает уровень ставок.

Рисунок 6.1 представляет собой срез для одного среза норматива CAR ( $CAR = 10\%$ ). Из Рисунок 6.1 следует, что существует такой норматив достаточности капитала, при котором синяя пунктирная линия пересечется со сплошной оранжевой линией. Изменение соотношения прибылей на традиционном и теневом секторе в зависимости от CAR можно увидеть на Рисунок

6.2. При заданных параметрах установление норматива свыше 46% приведет к переходу банков из традиционной в теневой банкинг.

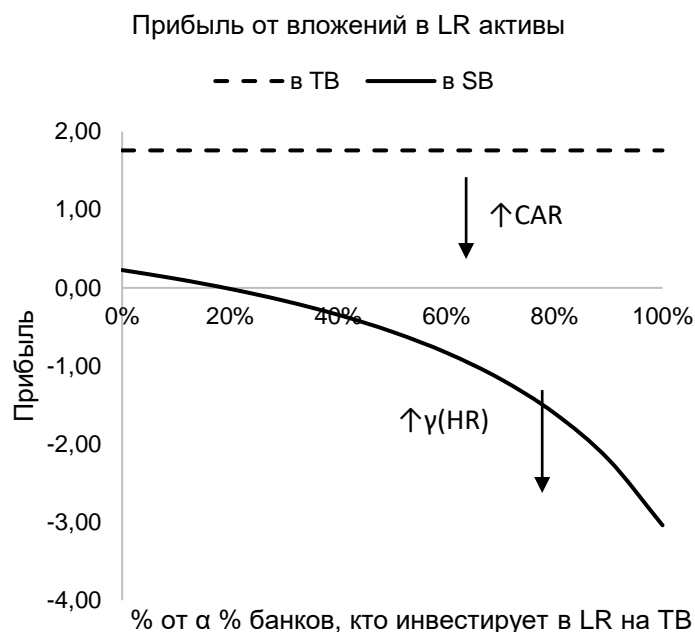


Рисунок 6.1 – Сравнение прибыли от вложений в низкорискованные активы (LR) в традиционном (TB) и теневом (SB) банкинге в зависимости от доли вложений в LR на TB (параметры для графика:  $\gamma(\text{LR-TB}) = 100\%$ ;  $\gamma(\text{HR-SB}) = 50\%$ ;  $\text{CAR} = 10\%$ ,  $\alpha = 40\%$ ,  $R(\text{NR-TB}) = 1.3$ ,  $R(\text{LR-TB}) = 1.5$ ,  $R(\text{HR-SB}) = 2.0$ , использовался вариант 1 формирования ставок инвесторами)

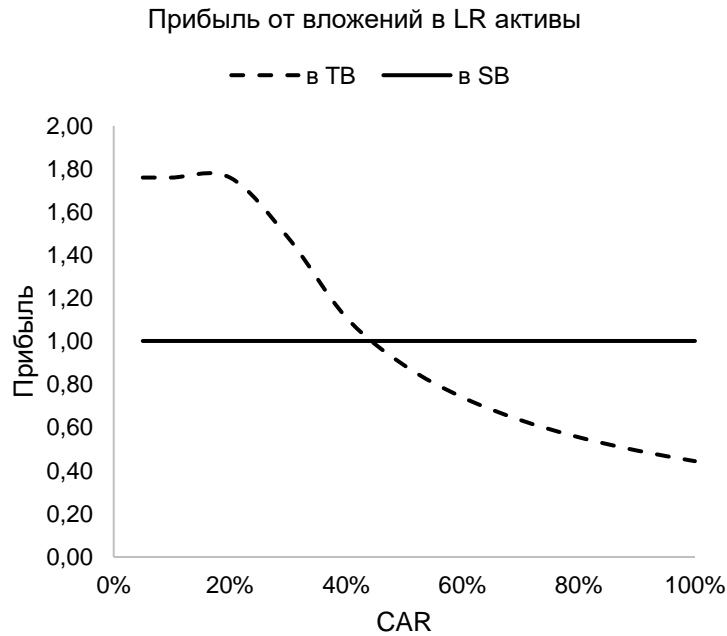


Рисунок 6.2 – Сравнение прибыли от вложений в низкорискованные активы (LR) в традиционном (TB) и теновом (SB) банкинге в зависимости от CAR (параметры для графика:  $\gamma(\text{LR-TB}) = 50\%$ ;  $\gamma(\text{HR-SB}) = 10\%$ ;  $\alpha = 40\%$ ;  $R(\text{NR-TB}) = 1.3$ ,  $R(\text{LR-TB}) = 1.5$ ,  $R(\text{HR-SB}) = 2.0$ )

Вариант 2 формирования ставок инвестора.

В отсутствии вложений в HR в теновом банкинге  $R_{rest}$  может оказаться меньше  $R_{LR-TB}$  за счет того, что в  $R_{rest}$  включается ценообразование по безрисковым активам. В этом случае прибыль от вложений в LR риски на SB всегда выше. Более того, разрыв между прибылью на SB и TB увеличивается с ростом CAR.

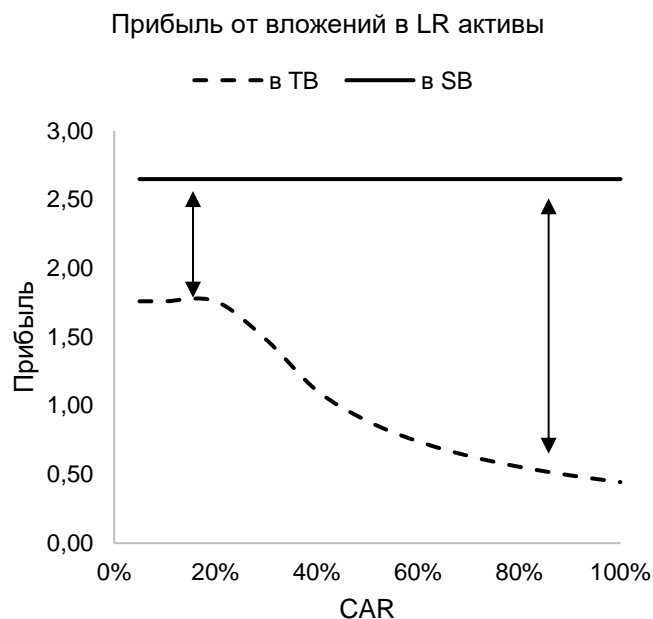


Рисунок 6.3 – Разрыв прибыли на SB и TB от вложений в LR при формировании ставки по методу 2 (в отсутствие вложений в HR на SB) (параметры для графика:  $\gamma(\text{LR-TB}) = 100\%$ ;  $\gamma(\text{HR-SB}) = 0\%$ ;  $\alpha = 40\%$ ,  $R(\text{NR-TB}) = 1.3$ ,  $R(\text{LR-TB}) = 1.5$ ,  $R(\text{HR-SB}) = 2.0$ , метод 2 формирования ставки инвесторами)

Однако с увеличением доли вложений в HR активы, ставка  $R_{rest}$  увеличивается. Найдем предельный уровень вложений в HR, при котором вкладываться в LR риски будет выгоднее на TB по сравнению с SB (рисунок 6.3). При  $x_{sb-inf} > \overline{x_{sb-inf}}$ , ставка инвестора  $R_{rest}$  настолько высока, что не покрывает выгоды от отсутствия ограничения на вложения из-за норматива. Таким образом, при  $x_{sb-inf} > \overline{x_{sb-inf}}$  традиционный (нетеневого) банкинг с регулированием будет предпочтительней теневого банкинга без регулирования для вложений в LR.

#### 6.1.4 Вычет потерь вкладчиков из благосостояния

Одним из спорных моментов в модели [305] является неучет потерь вкладчиков в расчете благосостояния общества. Далее показано, что вывод о положительном влиянии теневого банкинга может быть опровергнут, если в расчет благосостояния включать потери вкладчиков.



На Рисунок 6.4 на первом графике показан расчет благосостояния для LR активов, на втором графике – для HR активов. Если сравнить оранжевую линию с круглым маркером, то благосостояние в размере прибыли от вложений в HR активы выше. Однако если мы учтем потери вкладчиков, то соотношение прибылей для LR и HR активов изменится. Мы видим, что потери вкладчиков от вложений в HR активы сильно больше, что создает такой значимый разрыв между линиями на втором графике.

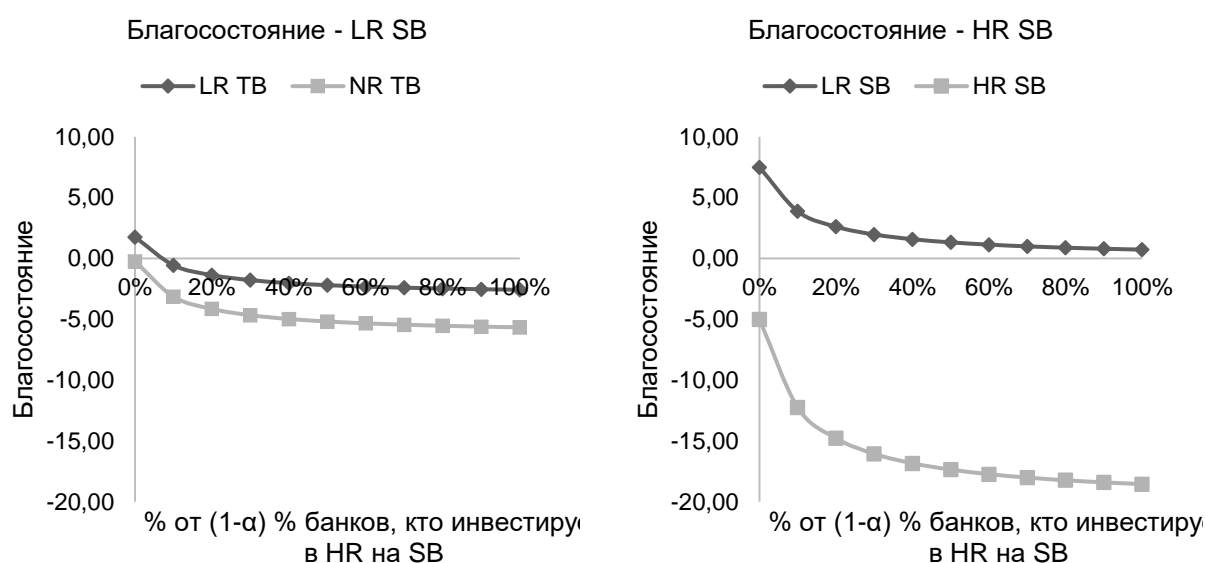


Рисунок 6.4 – Сравнение уровня благосостояния общества с учетом и без потерь инвесторов от участия банков во вложениях в традиционном и теневом банкинге (параметры для графика:  $\gamma(\text{LR-TB}) = 90\%$ ;  $\gamma(\text{HR-SB}) = 50\%$ ;  $\text{CAR}=30\%$ ,  $\alpha = 40\%$ ,  $R(\text{NR-TB}) = 1.3$ ,  $R(\text{LR-TB}) = 1.5$ ,  $R(\text{HR-SB}) = 2.0$ )

#### 6.1.5 Сравнение прибыли в традиционном и теневом банкинге

Внедрив три модификации в модель (регулирования капитала, режим ПВР и расчет благосостояния с учетом потерь), рассмотрим, как изменяются выводы о положительном влиянии теневого банкинга на благосостояние общества.

Пусть к LR активам имеют доступ только 40% банков. У каждого банка имеется капитал в размере 2 ед. и инвестиции в размере 20 ед. Иные характеристики представлены в Таблица 6.2.

Таблица 6.2 – Исходные параметры

Актив	Тип банкинга	Ожидаемые потери, EL	Процент по кредиту, R L	Коэффициент риска, RW
LR	TB	0.20	2.00	0.90
NR	TB	0.10	1.50	0.50
LR	SB	0.20	2.00	-
NR	SB	0.50	2.50	-

Перед началом анализа отметим, что показанные далее числовые выводы приведены не с целью создать бенчмарк, но показать на числовом (более простом) примере, что возможны ситуации, которые противоречат выводам [305]. Иначе говоря, что существуют случаи, когда традиционный банкинг приносит выше уровень благосостояния чем теневая даже при наличии регулирования капитала. То есть можно найти такое значение CAR, при котором прибыль банков в традиционном банкинге все еще выше, чем в теневом.

При формировании ставки вариантом 1.

Воздействие CAR на NR активы меньше чем на LR из-за более низкого риск-веса. Если бы был неограниченный объем NR и LR активов, то все средства  $(Inv + k)$  можно было бы вложить в NR активы при  $CAR = 18.18\%$ , в LR активы - при  $CAR = 10.10\%$ . CAR выше заданных значений приводит к ограничению инвестирования (возможно инвестировать, только определенную долю от  $Inv + k$ , где доля рассчитывается как  $k / ((Inv + k) * RW * CAR)$ ).

В рассматриваемой модели объем LR активов ограничен (в нашем примере только 40% банков имеют к ним доступ, или каждый банк может инвестировать не более 40% средств при равномерном распределении доступа). Следовательно, для LR активов можно найти такой CAR, который ограничивает вложения до 40%. Этот уровень CAR составляет 25.25%.

На Рисунок 6.5 показано, что традиционный (нетеневой) банкинг прибыльней и привлекательней для банков до  $CAR = 34\%$ . При CAR выше 34% для инвестирования в LR активы банки будут переходить на теневой банкинг. На Рисунок 6.6 показано, что если прибыль банков считать за вычетом потерь вкладчиков, то традиционный (нетеневой) банкинг приносит при CAR до 100% благосостояние обществу всегда выше.

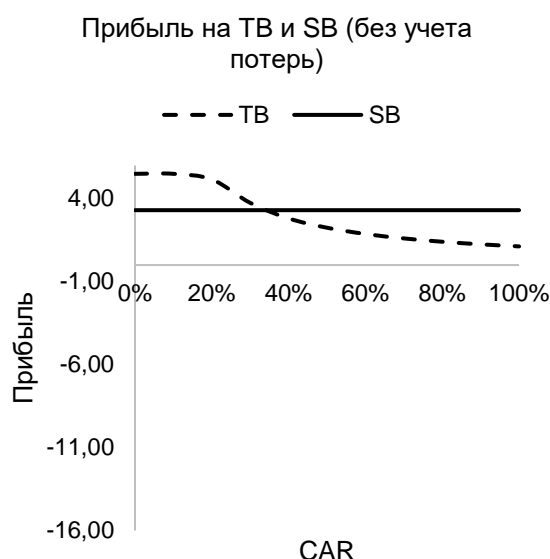


Рисунок 6.5 – Благополучие в размере прибыли банков без учета потерь вкладчиков

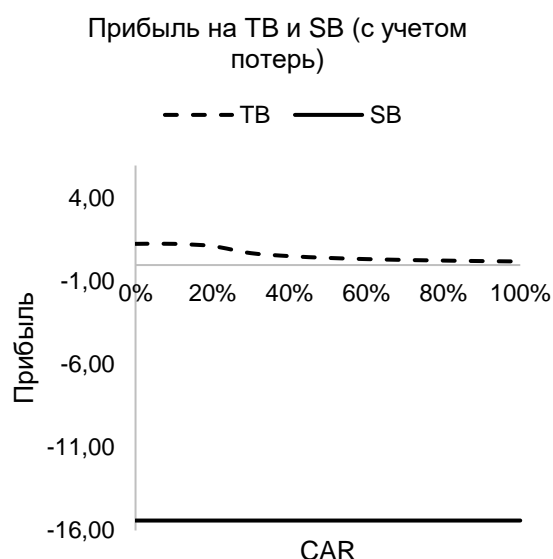


Рисунок 6.6 – Благополучие в размере прибыли банков при учете потерь вкладчиков

Можно сказать, что рисунки выше представляют два взгляда на систему: взгляд банка (Рисунок 6.5) и регулятора (Рисунок 6.6). При высоких потерях в теновом банкинге регулятор не может принять решение оставлять теновой банкинг (не проводить надзор банковской системы с целью выявления таких банков), так как видит существенное влияние этого банкинга на благополучие общества. Таким образом, вывод [305] о том, что теновой банкинг следует оставить, так как он приносит выше благополучие, оказывается неверным.

#### 6.1.6 Некоторые выводы

В [305] показано, что запрет тенового банкинга не выгоден для общества, так как прибыль, получаемая в теновом банкинге увеличивает благополучие общества. В настоящей работе данный вывод опровергнут и показано, что существуют случаи, когда традиционный (нетеновой) банкинг выгоден даже при наличии регулирования капитала.

В работе в модель [305] были добавлены режим ПВР для банков, регулирование капитала, учет потерь вкладчиков и взвешивание прибыли на риск непредвиденных потерь. Это было сделано с целью проверить, остается ли

устойчивым вывод о положительном влиянии прибыли в теневом банкинге на благосостояние общества.

Получено, что режим ПВР позволяет достичь первого наилучшего решения по [305]. Однако first-best действует при  $CAR < \overline{CAR}$ . В этом случае банки инвестируют в NR и LR активы в традиционном (нетеневом) банкинге. При усилении регулирования ( $CAR > \overline{CAR}$ ) first-best перетекает в случай, когда наряду с традиционным действует и теневой банкинг. Поскольку ПВР позволяет инвестировать в LR активы в рамках традиционного банкинга, и first-best обладает наиболее высоким уровнем благосостояния, то получается, что случай, когда действует только традиционный банкинг, приносит благосостояния больше чем теневой. При этом CAR хотя и ограничивает вложения в активы в традиционном (нетеневом) банкинге, но уменьшает объем ожидаемых потерь. Вывод о негативном влиянии высоких требований к капиталу совпадают с выводом [306], и не подтверждают вывод [308].

В исследовании показано, что благосостояние корректнее считать как прибыль за вычетом потерь вкладчиков. В этом случае, даже если прибыль в теневом банкинге выше, чем в традиционном, потери вкладчиков настолько высоки, что не покрывают выигрыш, полученный в теневом банкинге. Таким образом, для регулятора предложение [305] оставить теневой банкинг не может быть принято.

## **6.2 Модели специализации в абстрактных системах: случай структурных ограничений, задаваемых неравенствами**

### **6.2.1 Постановка задачи**

Системы, состоящие из конечного числа элементов, каждый из которых способен выполнять одну или несколько функций так, что выполнение одной из них ведет к снижению выполнения хотя бы одной другой, повсеместно встречаются в различных областях знания. Проблема возникновения функциональной специализации в системах описанного типа, которые мы будем называть многоклеточными, является фундаментальной. В частности, возник-

новение клеточной специализации – один из ключевых вопросов в эволюционной биологии, так как возникновение дифференцированных типов клеток ведет к структурному усложнению организмов и увеличению биологической сложности. В эволюционной антропологии зарождение разделения труда в первобытных сообществах является двигателем к развитию экономики. Вопрос специализации может возникать и на уровне системы университетов, когда решается проблема оптимального распределения задач между университетами. Таким образом, выявление факторов, влияющих на возникновение специализации в многоклеточных системах, и развитие аппарата для исследования данного феномена является актуальной задачей.

Мы изучаем модель специализации в абстрактных системах, обобщая понятия структурных ограничений элементов, задавая их в форме неравенств.

Рассмотрим абстрактную систему из  $N$  элементов. Пусть  $i = 1 \dots N$  – индексы элементов в системе. Предположим, что каждый элемент может выполнять одну из двух задач. Обозначим через  $b_i \in [0, b_i^{\max}]$  уровень выполнения первой задачи элементом с индексом  $i$ , через  $v_i \in [0, v_i^{\max}]$  – уровень выполнения второй задачи элементом с индексом  $i$ .

Теперь определим уровни выполнения задач всей системой целиком. Мы будем полагать, что уровень выполнения системой первой задачи ( $B$ ) есть аддитивная функция переменных  $b_i$ , а уровень выполнения системой второй задачи ( $V$ ) есть аддитивная функция переменных  $v_i$ :

$$B(b) = \sum_{i=1}^N b_i; \quad V(v) = \sum_{i=1}^N v_i.$$

Определим индекс эффективности системы ( $W$ ) следующим образом:

$$W = V^\alpha \cdot B^\beta.$$

Мы используем данный индекс, следуя работе [309], и исходя из того, что данный индекс отражает синергетический эффект, возникающий между элементами одной системы.

Будем предполагать, что внутренняя структура каждого элемента задается структурным ограничением элемента:

$$v_i \leq \varphi(b_i),$$

где функция  $\varphi$  обладает рядом свойств, описанных в [310].

Мы будем предполагать также, что внешняя среда оказывает влияние на эффективность функционирования системы. Пусть  $C$  – это уровень ресурса, доступный системе при заданных параметрах внешней среды. Предположим, что для выполнения единицы первой задачи системе требуется  $k_1$  единиц ресурса, а для выполнения единицы второй задачи –  $k_2$  единиц ресурса. Таким образом, общее ресурсное ограничение имеет вид:

$$k_1 \sum_{i=1}^N b_i + k_2 \sum_{i=1}^N v_i \leq C, \text{ где } k_1, k_2, C > 0.$$

Оптимальной стратегией системы мы будем называть набор  $(b^*, v^*)$  таких уровней выполнения задач каждым из элементов, которые при условии существования структурных ограничений и общего ресурсного ограничения позволяют максимизировать эффективность функционирования системы. Таким образом, мы можем свести нашу задачу к классической задаче математического программирования с ограничениями:

$$\left\{ \begin{array}{l} W = \left( \sum_{i=1}^N b_i \right)^\alpha \cdot \left( \sum_{i=1}^N v_i \right)^\beta \rightarrow \max_{b,v} \\ \forall i = 1..N: v_i \leq \varphi(b_i), \\ k_1 \sum_{i=1}^N b_i + k_2 \sum_{i=1}^N v_i \leq C, \\ \forall i = 1..N: v_i \geq 0, \\ \forall i = 1..N: b_i \geq 0. \end{array} \right.$$

### 6.2.2 Основные результаты

Заметим, что данная задача имеет непустое множество допустимых решений при любом допустимом наборе параметров, так как точка  $(0, 0)$  принадлежит множеству допустимых решений Задачи 1 при всех допустимых значениях параметров модели. Далее мы приведем основные результаты относительно решения Задачи 1, и сравним эти результаты с выводами, полученными при рассмотрении модели со структурными ограничениями, имеющими вид равенств [310]. Можно выделить три случая различной структуры решения Задачи 1:

Случай 1'. Предположим, что в модели нет ресурсного ограничения и определим в данной ситуации множество оптимальных стратегий системы,  $D$ . Очевидно, что для каждой оптимальной стратегии все структурные ограничения выполняются как равенства. Теперь добавим в систему ресурсное ограничение. Под случаем 1' предполагается ситуация, когда как минимум одна стратегия из множества  $D$  допустима, то есть, ресурсное ограничение не влияет на максимально достижимый системой уровень эффективности. Можно сделать вывод, что Случай 1' возникает при условии, что во внешней среде содержится достаточно большое количество ресурса.

Случай 2'. Данный случай подразумевает, что ресурсное ограничение влияет на оптимальный уровень функционирования системы. Рассмотрим следующее множество:

$$A' = \{(b, v) \in R^{2N} | b_i > 0, 0 \leq v_i \leq \varphi_i(b_i)\},$$
$$\sum_{i=1}^N b_i = \frac{c}{k_1 \cdot (1 + \frac{\beta}{\alpha})}, \sum_{i=1}^N v_i = \frac{c}{k_2 \cdot (1 + \frac{\alpha}{\beta})}.$$

Случай 2' подразумевает, что множество  $A'$  не пусто. Все стратегии из данного множества предписывают системе потреблять все доступное для нее количество ресурса. Также множество  $A'$  совпадает с множеством решений Задачи 1.

Случай 3'. Данный случай наблюдается тогда и только тогда, когда ресурсное ограничение оказывает влияние на максимально достижимый уровень индекса функционирования, и множество  $A'$  является пустым. В данном случае, поиск решения усложняется. В частности, если все функции компромиссов являются линейными или выпуклыми, оптимальные стратегии системы могут быть найдены следующим образом. Необходимо найти стратегии, которые максимизируют или минимизируют (зависит от параметров модели) значение общего уровня выполнения первой функции системой при ограничениях, задаваемых структурными ограничениями и при предположении, что все доступное количество ресурса полностью потребляется системой.

Взаимосвязь модели со структурными ограничениями в виде равенств и модели со структурными ограничениями в виде неравенств.

Рассмотрим Задачу 1 со структурными ограничениями в виде равенств (Задача 1а) и такую же задачу со структурными ограничениями в виде неравенств (Задача 1б). Нас будет интересовать, каким образом будет меняться структура решения при переходе от одной задачи к другой.

Предположим, что изначально мы имели Случай 1 Задачи 1а. Замена структурных ограничений в виде равенств на структурные ограничения в форме неравенств не изменят решения задачи. Более того, после указанных изменений, мы получим Случай 1' Задачи 1б. Предположим теперь, что изначально у нас имелся Случай 2 решения Задачи 1а. После замены всех структурных ограничений в виде равенств на структурные ограничения в форме неравенств мы получим Случай 2' Задачи 1б, и все решения Задачи 1а станут решениями Задачи 1б. Предположим, что изначально у нас имелся Случай 3 решения Задачи 1а. После замены равенств на неравенства в структурных могут возникнуть две различных ситуации. В первой ситуации получаем Случай 2' Задачи 1б. В данном случае все оптимальные стратегии предполагают существование хотя бы одного такого элемента в системе, что этот элемент будет иметь уровень выполнения второй функции меньше того, который пред-



писывается функцией компромиссов этого элемента при заданном уровне выполнения первой функции. Во второй ситуации получаем Случай 3' Задачи 1б. Чтобы завершить анализ изменения структуры решения рассматриваемых задач, мы должны рассмотреть единственный оставшийся неисследованным случай, когда множество допустимых решений Задачи 1а пусто (Случай 0). Предположим, что изначально мы имели Случай 0 Задачи 1а. После замены структурных ограничений в виде равенств на структурные ограничения в форме неравенств, мы получим Случай 2' соответствующей Задачи 1б.

Отметим, что ситуация, когда рассматриваемая оптимизационная задача имеет пустое множество допустимых решений, возникает только в случае структурных ограничений, заданных равенствами. Данный случай (Случай 0) возникает исключительно тогда, когда количество ресурса, доступного системе, мало. Иными словами, существует некоторое пороговое значение ресурса, такое, что если уровень концентрации ресурса во внешней среде системы опустится ниже порогового значения, система перестанет функционировать.

Итого, мы показали, что замена структурных ограничений в форме равенств на структурные ограничения в виде неравенств приводит к некоторым изменениям в структуре решения в Случаях 0 и 3 задачи математического программирования со структурными ограничениями в виде неравенств.

Зависимость структуры решения модели со структурными ограничениями в виде неравенств от количества доступного системе ресурса.

Рассмотрим Задачу 1. Тогда существуют допустимые значения количества ресурса, содержащегося во внешней среде,  $C_1$ ,  $C_2$ ,  $C_1 \geq C_2$ , зависящие от прочих значений параметров модели и такие, что: (1)  $\forall C \geq C_1$  наблюдается Случай 1' решения указанной задачи; (2)  $\forall C < C_2$  реализуется Случай 2' решения указанной задачи.

Иными словами, в Задаче 1 Случай 1' возникает в ситуации среды, богатой на ресурс; Случай 2' возникает в ситуации среды, бедной на ресурс; Случай 3', если возникает, наблюдается в ситуациях, когда количество ресурса,

содержащегося во внешней среде, принимает промежуточные значения. Такая явная зависимость структуры решения от количества допустимого ресурса не наблюдается в модели со структурными ограничениями в форме равенств.

### 6.2.3 Некоторые выводы

1) Сформулирована и исследована модификация модели специализации в многоклеточных системах, предложенной в работе [310]. Данная модификация заключается в обобщении структурных ограничений элементов. Предлагается рассматривать структурные ограничения не в виде жестких функциональных зависимостей между уровнями выполнения функций элементами, а в виде неравенств, когда при заданном фиксированном уровне выполнения одной функции элемент может иметь уровень выполнения другой функции, меньше или равный некоторого значения, зависящего от уровня выполнения первой функции.

2) Получены результаты о структуре решения поставленной задачи математического программирования со структурными ограничениями в виде неравенств. Произведено сравнение различных случаев решения задачи с ограничениями в виде неравенств и решений соответствующей задаче со структурными ограничениями, задаваемыми в виде равенств. Показано, каким образом изменится структура решения поставленной задачи при переходе от структурных ограничений в виде равенств к структурными ограничениям, задаваемым неравенствами. Получены теоретические результаты, позволяющие выявить, каким образом изменяется структура решения в модели со структурными ограничениями в виде неравенств в зависимости от изменения количества доступного системе ресурса.

## **6.3 Разработка матричного метода линеаризации систем уравнений в частных производных на постоянных решениях**

Нелинейные уравнения в частных производных часто встречаются в различных областях математики, физики и гидродинамики и других наук. Их изучение часто начинают с линеаризации на постоянных решениях, если таковые существуют, затраты времени на которую могут быть достаточно большими.

Общепринятым подходом к данной проблеме является запись решения в нормализованном виде относительно малых возмущений и подстановка его в систему. Затем все слагаемые второго и выше порядков малости отбрасываются, и получается линеаризованная система уравнений. Тем не менее, для достаточно большой системы (например, 3 D система уравнений газовой динамики пяти уравнений с пятью неизвестными), необходимо принять во внимание много слагаемых и тщательно их вычислить с соответствующими временными затратами. В настоящем исследовании был предложен алгоритм, который помогает автоматизировать часть этой работы. Новый подход существенно упрощает сложные вычисления с помощью использования специального объекта.

Этот алгоритм основан на построении кольца функциональных матриц специального вида. Каждое уравнение системы преобразовывается в соответствующее матричное уравнение, где единственным ненулевым элементом всех входящих в него матриц является правый верхний элемент. Приравнивая его к нулю, получим соответствующее линеаризованное уравнение системы. По теореме о гомоморфизме колец [311], выполняя матричные операции в кольце, мы не различаем никаких слагаемых второго и выше порядка малости и, тем самым, процесс линеаризации убирает необходимость в кропотливой проверке каждого такого слагаемого в каждом уравнении системы.

Приведен пример такой линеаризации, применённой к 3D уравнениям газовой динамики с квазигидродинамической регуляризацией, которая состоит из пяти уравнений с пятью неизвестными [312].

## ЗАКЛЮЧЕНИЕ

При выполнении проекта получены следующие основные результаты, могут быть сделаны следующие выводы и рекомендации:

- разработаны и исследованы новые средние значения выборок в рамках теории важности критериев, основанные на идее многокритериальной оптимизации;

- в рамках продолжающегося исследования по разработке и анализу моделей расчета степени манипулируемости правил агрегирования предпочтений произведены расчеты, показывающие, какая процедура агрегирования будет наименее манипулируемой для различных параметров (моделей голосования, расширения предпочтений, числа альтернатив, числа участников); разработано соответствующее программное обеспечение и найдены оценки вычислительной сложности алгоритмов; найдены условия, когда манипулирование всегда безопасно и когда существует небезопасно манипулируемый профиль для двухступенчатой мажоритарной системы и для правила Коупленда;

- в рамках исследования профилей предпочтений найдены вероятности однопиковых профилей предпочтений для разных моделей генерирования таких профилей;

- найдены достаточные условия непустоты множества обобщенных турнирных решений, основанных на принципе устойчивости; показано, что эти же условия будут справедливы для некоторых турнирных решений и в случае бесконечного множества вариантов;

- применены методы анализа латентных классов и дисперсионного анализа к обработке мониторинговых данных, выделены и проанализированы различные группы студентов по их отношению к переводу занятий в онлайн формат;

- построено обобщение разработанных ранее индексов дальних взаимодействий (LRIC), учитывающие взаимозависимость между элементами сети; эти индексы применены к анализу сетей международной торговли и взаимодействия террористических групп;

- разработаны индексы центральности, учитывающие индивидуальные и групповые влияния вершин сети; эти индексы применены к анализу сетей международной миграции и межстрановой торговли;
- разработана модель по оценке схожести слоёв многослойной динамической сетевой структуры; эта модель применена к анализу сети международного экспорта/импорта продовольствия;
- рассмотрена модификация гравитационной модели, включающей в себя индексы центральности сети; эта модель применена к анализу многосторонних торговых потоков;
- разработан и исследован способ оценивания центральностей многослойной сети методами теории функций доверия; доказаны взаимосвязи полученных оценок с агрегированием классических послойных центральностей в некоторых частных случаях;
- разработан гибридный метод экономного подъема нечеткого тематического кластера в таксономии предметной области; этот метод применен к выявлению исследовательских тенденций по анализу статей, опубликованных в журналах издательств Springer и Elsevier по тематике Наука о Данных;
- разработан и исследован метод альтернирующих дихотомий для кластеризации данных; этот метод применен для анализа фондового рынка;
- при решении задачи оценивания экономической целесообразности создания нового грузового транспортного узла сформулировано и доказано достаточное условие конкурентоспособности тарифов на услуги в таком узле; сформулированы все рассматриваемые в нем прикладные задачи и доказаны соответствующие утверждения;
- продемонстрирована возможность решения задачи об оценке финансовой целесообразности строительства и эксплуатации нового транспортного узла, как задачи нахождения максимума максимина разности двух билинейных функций;

- предложена модификация ранее разработанной модели оптимального размещения зарядных станций автомобилей вдоль участка шоссе для решения аналогичной задачи размещения городских сетей таких станций.
- проведен анализ изменение населения в Арктике из-за естественного прироста и чистой миграции населения;
- проанализированы существующие методы распределения спорных территорий и предложен ряд новых модификаций этих методов; эти методы апробированы на примере раздела спорных территорий стран Арктического региона;
- проанализировано влияние изменения климата на социально-экономическое развитие населения Арктики;
- проведен анализ климатических и географических данных стран Африки относительно влияния этих факторов на водообеспеченность и водопотребление странами;
- сделан аналитический обзор научной литературы по исследованию влияния изменения климата на продовольственную безопасность стран Африки;
- разработаны интервальные модификации оболочечного анализа данных; эти методы апробированы на решении задачи оценки эффективности превентивных противопожарных мероприятий в регионах РФ;
- проанализирована и модифицирована популярная модель образования теневой банковской системы; доказано, что политика регулятора не может основываться на поддержке банковского теневого сектора;
- разработана и исследована модификация модели специализации в многоклеточных системах в случае структурных ограничений, задаваемых неравенствами; произведено сравнение решений задачи нахождения стратегии, максимизирующей эффективность функционирования системы, для разных случаев ограничений;

– предложен матричный метод линеаризации систем уравнений в частных производных на постоянных решениях, позволяющий автоматически получить линеаризованную систему, подавая на вход лишь исходную нелинеаризованную систему.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Отчет МНУЛ анализа и выбора решений, ТЗ-37 за 2019г.
2. Средние величины. [Электронный ресурс]. – 2006. – Grandars.ru: <http://www.grandars.ru/student/statistika/srednie-velichiny.html> (дата обращения 19.10.2020).
3. Бурханова И.И. Теория статистики. – 2009. – ВикиЧтение: [Электронный ресурс]. <https://econ.wikireading.ru/33154> (дата обращения 19.10.2020).
4. Джини К. Средние величины. – М.: Статистика, 1970. – 447 с.
5. Пфанцагль И. Теория измерений. – М.: Мир, 1976. – 248 с.
6. Орлов А.И. Связь между средними величинами и допустимыми преобразованиями шкалы // Математические заметки. – 1981. – Т. 30. – С. 561 – 568.
7. Подиновский В.В., Ногин В.Д. Парето-оптимальные решения многокритериальных задач. – М.: Физматлит, 1982. – 248 с. Изд. второе, испр. и доп. – М.: Физматлит, 2007. – 260 с.
8. Подиновский В.В. Многокритериальные задачи с однородными равноценными критериями // Журнал вычислительной математики и математической физики. – 1975. – Т. 15. – С. 330 – 344.
9. Fishburn, P.C. Decision and Value Theory. – New York: Wiley, 1964. – 437 p.
10. Маршалл А., Олкин И. Неравенства: теория мажоризации и ее приложения. – М.: Мир, 1983. – 574 с.
11. Подиновский В.В. Количественная важность критериев с непрерывной шкалой первой порядковой метрики // Автоматика и телемеханика. – 2005. – № 9. – С. 129 – 137.
12. Кнут Д.Э. Искусство программирования, т. 3. Сортировка и поиск, 2-е изд. – М.: Вильямс, 2007. – 822 с.



13. Шоломицкий, А.Г. Теория риска. Выбор при неопределенности и моделирование риска. – М.: Изд. дом ГУ ВШЭ, 2005. – 401 с.
14. Льюс Р.Д., Райфа Х. Игры и решения. Введение и критический обзор. – М.: Изд-во иностр. лит-ры, 1961. – 642 с.
15. Подиновский В.В., Нелюбин А.П. Средние величины: многокритериальный подход // Проблемы управления. – 2020. – № 5. – С. 3 – 16.
16. Gibbard A. Manipulation of voting schemes// *Econometrica*. –1973. – N 41. P. 587–601.
17. Satterthwaite M. Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions// *Journal of Economic Theory*. – 1975. – N10. –P. 187–217.
18. Chamberlin J.R. An investigation into the relative manipulability of four voting systems// *Behavioral Science*. –1985. – N 30(4). – P.195–203.
19. Aleskerov F., Kurbanov E. Degree of manipulability of social choice procedures // Alkan A. et al. (eds.) *Current Trends in Economics. Studies in Economic Theory*. Berlin Heidelberg, N.Y.: – Springer, 1999. – N 8. – P.13–27.
20. Aleskerov F., Karabekyan D., Sanver R., Yakuba V. On manipulability of positional voting rules// *SERIEs: Journal of the Spanish Economic Association*. – 2011. – N2 (4). – P. 431–446.
21. Aleskerov F., Karabekyan D., Sanver R., Yakuba V. On the manipulability of voting rules: Case of 4 and 5 Alternatives// *Mathematical Social Sciences*. – 2012. – N 64 (1). – P. 67–73.
22. Pritchard G., Wilson M. Exact results on manipulability of positional voting rules// *Social Choice and Welfare*. – 2007. – N 29. – P. 487–513.
23. Favardin P., Lepelley D. Some further results on the manipulability of social choice rules// *Social Choice and Welfare*. – 2006. – N 26. – P. 485–509.
24. Nitzan S. The vulnerability of point-voting schemes to preference variation and strategic manipulation// *Public Choice*. – 1985. – N 47. – P. 349–370.
25. Kelly J. Almost all social choice rules are highly manipulable, but few aren't// *Social Choice and Welfare*. – 1993. – N 10. – P. 161–175.

26. Карабекян Д.С. Манипулирование в задаче коллективного принятия решений// диссертация на соискание ученой степени кандидата экономических наук. – Москва, 2012. 171с.
27. Веселова Ю.А. Всегда ли безопасно коалиционное манипулирование результатом голосования?// В кн.: Управление развитием крупномасштабных систем (MLSD'2019): материалы Двенадцатой международной конференции, 1–3 окт. 2019 г. / Под общ. ред.: С. Н. Васильев, А. Цвиркун. Институт проблем управления им. В.А. Трапезникова РАН. – 2019. – С. 1185–1187.
28. Иванов А.А. Эффективные вычислительные схемы расчета манипулируемости процедур агрегирования// Информационные технологии и вычислительные системы. – 2020. – N 2. – С. 38–50.
29. Veselova Y.A. Does Incomplete Information Reduce Manipulability?// Group Decision and Negotiation. – 2020. – N 29(3). – P. 523–548.
30. Arrow K.J. Social Choice and Individual Values. – New York: Wiley, 1951.
31. May K.O. A set of independent necessary and sufficient conditions for simple majority decisions// Econometrica. – 1952. N 20(4). – P. 680 – 684.
32. Aşan G., Sanver M.R. Another characterization of the majority rule// Econ. Lett. – 2002. – N 75(3). – P. 409 – 413.
33. Campbell D.E., Kelly J.S. A simple characterization of majority rule// Econ. Theory. – 2000. – N 15(3). – P. 689 – 700.
34. Woeginger G.J. A new characterization of the majority rule// Econ. Lett. – 2003. – N 81(1). – P. 89 – 94.
35. Inada K. A Note on the Simple Majority Decision Rule// Econometrica. – 1964. – N 32(4). – P. 525 – 531.
36. Inada K. The Simple Majority Decision Rule// Econometrica. – 1969. – N 37(3). – P. 490 – 506.
37. Sen A.K. A Possibility Theorem on Majority Decisions// Econometrica. – 1966. – N 34(2). – P. 491 – 499.

38. Black D. On the rationale of group decision-making// J. Polit. Econ. 1984. – N 56. P. 23 – 34.
39. Brandt F., Brill M., Hemaspaandra E., Hemaspaandra L.A. Bypassing combinatorial protections: Polynomial-time algorithms for single-peaked electorates// J. Artif. Intell. Res. – 2015. – N 53. – P. 439 – 496.
40. Puppe C. The single-peaked domain revisited: A simple global characterization// J. Econ. Theory. – 2018. – N 176. – P. 55 – 80.
41. Gehrlein W.V., Lepelley D. Voting paradoxes and their probabilities// In Voting Paradoxes and Group Coherence, Springer-Verlag Berlin Heidelberg - 2011.- P. 1 – 47.
42. Wilson M.C., Pritchard G. Probability calculations under the IAC hypothesis// Math. Soc. Sci. – 2007. – N 54(3). – P. 244 – 256.
43. Lackner M.L., Lackner M. On the likelihood of single-peaked preferences// Soc. Choice Welf. – 2017. – N 48(4). – P. 717 – 745.
44. Karpov A. On the number of group-separable preference profiles// Group Decis. Negot. – 2019. – N 28(3). – P. 501 – 517.
45. Ballester M.A., Haeringer G. A characterization of the single-peaked domain// Soc. Choice Welf. – 2011. – N 36. – P. 305 – 322.
46. Karpov A. The likelihood of single-peaked preferences under classic and new probability distribution assumptions// Soc. Choice Welf. – 2020. – N 55. – P. 629–644.
47. Gehrlein W.V., Fishburn P.C. Condorcet's paradox and anonymous preference profiles// Public Choice. – 1976. – N 26. – P. 1 – 18.
48. Kuga K., Nagatani H. Voter antagonism and the paradox of voting// Econometrica. – 1974. – N 42. – P. 1045 – 1067.
49. Gehrlein W.V. The expected probability of Condorcet's paradox// Econ. Lett. – 1981. – N 7. – P. 33 – 37.
50. Karpov A. Preference Diversity Orderings// Group Decis. Negot. – 2017. – N 26(4). – P. 753 – 774.

51. Faliszewski P., Hemaspaandra E., Hemaspaandra L.A. The complexity of manipulative attacks in nearly single-peaked electorates// *Artif. Intell.* – 2014. – N 207. – P. 69 – 99.
52. Brederbeck R., Chen J., Woeginger G.J. Are there any nicely structured preference profiles nearby?// *Math. Soc. Sci.* – 2016. – N 79. – P. 61 – 73.
53. Laslier J.F. *Tournament Solutions and Majority Voting.* – Berlin: Springer - 1997. – P. 1-256.
54. Condorcet, Marquis de. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix.* – Paris: L'imprimerie royale - 1785. –P.1-195.
55. Duggan J. General conditions for the existence of maximal elements via the uncovered set// *Journal of Mathematical Economics.* – 2011. – N 47(6). – P. 755 – 759.
56. von Neumann J., Morgenstern O. *Theory of Games and Economic Behavior.* – Princeton: Princeton University Press -1944.-P. 1-776.
57. Aleskerov F., Subochev A. Matrix-vector representation of various solution concepts// Working paper WP7/2009/03. Moscow: SU – Higher School of Economics. 2009. - URL: <https://publications.hse.ru/mirror/pubs/share/folder/veuqm6ur2s/direct/133996732.pdf> (дата обращения 2020-08-11)
58. Wuffle A., Feld S.L., Owen G., Grofman B. Finagle's Law and the Finagle Point, a New Solution Concept for Two-Candidate Competition in Spatial Voting Games Without a Core// *American Journal of Political Science.* – 1989. – N 33(2). – P. 348 – 375.
59. Subochev A. Dominant, Weakly Stable, Uncovered Sets: Properties and Extensions: Working paper WP7/2008/03. Moscow: SU – Higher School of Economics, 2008. – URL: [https://www.hse.ru/data/2010/05/05/1216436938/WP7\\_2008\\_03-f.pdf](https://www.hse.ru/data/2010/05/05/1216436938/WP7_2008_03-f.pdf) (дата обращения 2020-08-11)
60. Aleskerov F., Subochev A. Modeling optimal social choice: matrix-vector representation of various solution concepts based on majority rule// *Journal of Global Optimization.* – 2013. – N 56(2). – P. 737 – 756.

61. Субочев А.Н. О возможности имплементации такой функции коллективного выбора, как объединение минимальных внешнеустойчивых множеств, и о других ее полезных свойствах и возможностях применения// В кн.: XVII Апрельская международная научная конференция по проблемам развития экономики и общества: в 4 кн. / Отв. ред.: Е. Г. Ясин. Кн. 1. – М.: Издательский дом НИУ ВШЭ. – 2017. – С. 111 – 120.
62. Miller N. A new solution set for tournaments and majority voting: Further graph-theoretical approaches to the theory of voting// American Journal of Political Science. – 1980. – N 24. – P. 68 – 96.
63. Fishburn P.C. Condorcet Social Choice Functions// SIAM Journal on Applied Mathematics. – 1977. – N 33(3). – P. 469 – 489.
64. Banks J.S., Duggan J., Le Breton M. Bounds for mixed strategy equilibria in the spatial model of elections// Journal of Economic Theory. – 2002. – N 103. – P. 88 – 105.
65. Banks J.S., Duggan J., Le Breton M. Social choice and electoral competition in the general spatial model// Journal of Economic Theory. – 2006. – N 126(1). – P. 194 – 234.
66. Egorova L.G., Torop V. Analysis of Students' Attitude towards Online Education// in New Perspectives in Science Education 9th Edition – International Conference (Florence, Italy, 19-20 March 2020). – Filodiritto Publisher. – 2020. – P. 112-116.
67. Freeman L.C. Centrality in social networks: conceptual clarification// Social Networks. – 1979. – N 1. – P. 215 – 239.
68. Bonacich P., Lloyd P. Eigenvector-like measures of centrality for asymmetric relations// Social Networks. – 2001. – N 23. – P. 191 – 201.
69. Bonacich P. Power and centrality: A family of measures// American journal of sociology. – 1987. – N 92(5). – P.1170 – 1182.
70. Katz L. A New Status Index Derived from Sociometric Index// Psychometrika. – 1953. – P. 39 – 43.

71. Brin S., Page, L. The anatomy of a largescale hypertextual Web search engine// Comput. Netw. – 1998. – N 30. – P. 107 – 117.
72. Kleinberg J.M. Authoritative sources in a hyperlinked environment// Journal of the ACM. – 1999. – N 46(5). – P. 604 – 632.
73. Rochat Y. Closeness Centrality Extended to Unconnected Graphs: The Harmonic Centrality Index// ASNA. – 2009. – URL: <https://infoscience.epfl.ch/record/200525> (дата обращения 2020-09-05)
74. Freeman L.C. A set of measures of centrality based upon betweenness// Sociometry. – 1977. – N 40. – P. 35 – 41.
75. Agneessens F., Borgatti S.P., Everett M.G. Geodesic based centrality: Unifying the local and the global// Social Networks. – 2017. – N 49. – P. 12 – 26.
76. Shimbel A. Structural parameters of communication networks// Bulletin of Mathematical Biology. – 1953. – N 15. – P. 501 – 507.
77. Everett M.G., Borgatti S.P. The centrality of groups and classes. Journal of Mathematical Sociology. – 1999. – N 23(3). – P. 181 – 201.
78. Piraveenan M., Prokopenko M., Hossain L. Percolation Centrality: Quantifying Graph-Theoretic Impact of Nodes during Percolation in Networks// PLoS ONE. – 2013. – N 8(1). – e53095.
79. Myerson R.B. Graphs and cooperation games// Math. Oper. Res. – 1977. – N 2. – P. 225 – 229.
80. Shapley L.S., Shubik M. A method for evaluating the distribution of power in a committee system// American Political Science Review. – 1954. – N 48. – P. 787 – 792.
81. Newman M.E.J. Mixing patterns in networks// Physical Review E. – 2003. – N 67(2) – P. 1-14.
82. Aleskerov F.T., Andrievskaya I.K., Permjakova E.E. Key borrowers detected by the intensities of their shortrange interactions// Working papers by NRU Higher School of Economics. Series FE "Financial Economics". – 2014. – No. WP BRP 33/FE/2014 – P. 1-18.

83. Aleskerov F., Meshcheryakova N., Shvydun S. Centrality measures in networks based on nodes attributes, long-range interactions and group influence// Cornell University. – 2016 – P.1- 44.
84. Aleskerov F.T., Meshcheryakova N.G., Shvydun S.V. Power in Network Structures// Models, Algorithms, and Technologies for Network Analysis. Springer Proceedings in Mathematics & Statistics. – 2017. – N 197. – P. 79 – 85.
85. Meshcheryakova N., Shvydun S. Power in Network Structures Based on Simulations// Complex Networks & Their Applications VI. COMPLEX NETWORKS 2017. Studies in Computational Intelligence. Springer, Cham. – 2018. – N 689. –1288p.
86. Shvydun S. Computational Complexity of SRIC and LRIC Indices// Network Algorithms, Data Mining, and Applications. NET 2018. Springer Proceedings in Mathematics & Statistics. Springer, Cham. – 2020. – N 315. – P. 1-244.
87. Aleskerov F., Yakuba V. Matrix-vector approach to construct generalized centrality indices in networks// WP7/2020/01. Moscow: Higher School of Economics Publ. House. – 2020 – P. 1-21.
88. Statistical Review of World Energy – URL: <https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html> (дата обращения 2020-04-01)
89. Pettersson T. UCDP/PRIO Armed Conflict Dataset Codebook v 20.1. – URL: <https://ucdp.uu.se/downloads/> (дата обращения 2020-11-07)
90. Pettersson T. UCDP Battle-related Deaths Dataset Codebook v 20.1.– URL: <https://ucdp.uu.se/downloads/> (дата обращения 2020-11-07).
91. Kivela M., Arenas A., Barthelemy M., Gleeson J.P., Moreno Y., Porter M.A. Multilayer networks// Journal of Complex Networks. – 2014. – N 2(3). – P. 203 – 271.
92. Holme P., Saramaki J. Temporal networks// Physics Reports. – 2012. – N 519(3) – P. 97 – 125.
93. De Domenico M., Nicosia V., Arenas A., Latora V. Structural reducibility of multilayer networks// Nature Communications. – 2015. – N 6(1). – P. 1- 9.

94. Nemeth R., Smith D. International Trade and World-System Structure: A Multiple Network Analysis// Review (Fernand Braudel Center). – 1985. – N 8(4). – P. 517 – 560.
95. Smith D.A., White D.R. Structure and Dynamics of the Global Economy: Network Analysis of International Trade 1965-1980// Social Forces. – 1992. – N 70(4). – P. 857 – 893.
96. Aleskerov F.T., Meshcheryakova N.G., Sergeeva Z.V., Shvydun S.V. Centrality Measures And Clustering Analysis in a Retail Food Network// 2017 IEEE 11th International Conference on Application of Information and Communication Technologies, vol. 1. M.: Institute of Electrical and Electronics Engineers. – 2017. – P. 48 – 52.
97. Meshcheryakova N. The Impact of Indirect Connections: The Case of Food Security Problem// Complex Networks and Their Applications VII. Studies in Computational Intelligence, vol. 813, Springer, Cham. – 2019. – P. 80 – 90.
98. Faust K. Comparison of methods for positional analysis: Structural and general equivalences// Social Networks. – 1998. – N 10(4). – P. 313 – 341.
99. Leicht E.A., Holme P., Newman M.E.J. Vertex similarity in networks// Physical Review E. – 2006. – N 73(2). – P.1-10.
100. Xulvi-Brunet R. Sokolov I.M. Changing correlations in networks: assortativity and dissortativity// Acta Phys. – 2005. – Pol. B 36. – P. 1431 – 1455.
101. Messmer B.T., Bunke H. Efficient subgraph isomorphism detection: a decomposition approach// IEEE Transactions on Knowledge and Data Engineering. – 2000. – N 12(2). – P. 307 – 323.
102. Raymond J.W. RASCAL: Calculation of Graph Similarity using Maximum Common Edge Subgraphs// The Computer Journal. – 2002. – N 45(6). – P. 631 – 644.
103. Wilson R.C., Zhu P. A study of graph spectra for comparing graphs and trees// Pattern Recognition. – 2008. – N 41(9). – P. 2833 – 2841.



104. Aleskerov F., Shvydun S. Stability and Similarity in Networks Based on Topology and Nodes Importance// Complex Networks and Their Applications VII. – 2018. – P. 94 – 103.
105. Tinbergen J. An Analysis of World Trade Flows// in Shaping the World Economy, edited by Jan Tinbergen. New York, NY: Twentieth Century Fund, - 1962. – P. 27-30
106. Anderson J.E. A theoretical foundation for the gravity equation // The American Economic Review. – 1979. – N 1(69). – P 106-116.
107. Anderson J.E, Van Wincoop E. Gravity with gravitas: A solution to the border puzzle: American economic review. - 2003. – P. 170-192.
108. Anderson J.E., Van Wincoop E. Trade costs// Journal of Economic literature. – 2004. – N 3(42). – P 691-751.
109. Baier S.L., Bergstrand J.H. Bonus vetus OLS: A simple method for approximating international trade-cost effects using the gravity equation// Journal of International Economics. – 2009. – N 1(77). – P 77-85.
110. Benedictis L., Tajoli L. The world trade network// The World Economy: Wiley Online Library. – 2011. – N 8(34). – P. 1417-1454
111. Benedictis L., Nenci S., Santoni G., Tajoli L., Vicarelli C. Network Analysis of World Trade using the BACI-CEPII// Global Economy Journal. – 2014. – N 4(14). – P. 287-343.
112. OECD Bilateral Trade in Goods by Industry and End-use (BTDIxE), ISIC Rev.4 . – URL: [https://stats.oecd.org/Index.aspx?DataSetCode=BTDIXE\\_I4](https://stats.oecd.org/Index.aspx?DataSetCode=BTDIXE_I4). (дата обращения 2020-10-10)
113. Shvydun S. Power of Nodes Based on Their Interdependence// Complex Networks XI. Springer Proceedings in Complexity. Springer, Cham. – 2020. – P. 70 – 82.
114. Global Terrorism Database. START: National Consortium for the Study of Terrorism and Responses to Terrorism. – URL: <http://start.umd.edu/gtd/> (дата обращения 2020-10-10)

115. Aleskerov F., Gavrilenkova I., Shvydun S., Yakuba V. Power Distribution in the Networks of Terrorist Groups: 2001–2018// Group Decis Negot. – 2020. – N 29. – P.399 – 424.
116. Eurostat, Migration and Migrant population Statistics. Eurostat. [Электронный ресурс] – 2019 – URL: [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Migration\\_and\\_migrant\\_population\\_statistics](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Migration_and_migrant_population_statistics). (дата обращения 2020-11-09).
117. OECD. International Migration Database. – 2019 – URL: <https://stats.oecd.org/Index.aspx?DataSetCode=MIG> (дата обращения 2020-10-10)
118. World Economic Situation Prospects. United Nations . –URL: [https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/WESP2019\\_BOOK-ANNEX-en.pdf](https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/WESP2019_BOOK-ANNEX-en.pdf) (дата обращения 2020-05-10)
119. Meshcheryakova N. Similarity Analysis in Multilayer Temporal Food Trade Network// Complex Networks XI. Proceedings of the 11th Conference on Complex Networks CompleNet 2020. Springer. – 2020. – P. 322 – 333.
120. WITS – About WITS 2019. – URL: [https://wits.worldbank.org/about\\_wits.html](https://wits.worldbank.org/about_wits.html). (дата обращения 2019-11-10)
121. Ranking of countries by commodity exports URL: [http://www.fao.org/faostat/en/#rankings/countries\\_by\\_commodity\\_exports](http://www.fao.org/faostat/en/#rankings/countries_by_commodity_exports) (дата обращения 2019-11-10)
122. Gaulier G., Zignago S. Baci: international trade database at the product-level (the 1994-2007 version). – 2010. – P. 1-35
123. Newman M.E.J. The Structure and Function of Complex Networks // SIAM Review, N 45(2), - 2003, P. 167–256.
124. Dempster A.P. Upper and Lower Probabilities Induced by a Multivalued Mapping// The Annals of Mathematical Statistics. – 1967. – N 38(2). P. 325 – 339.
125. Shafer G. A Mathematical Theory of Evidence. – Princeton University Press.- 1976.- P. 1-314

126. Gao C., Wei D., Hub Y., Mahadevan S., Deng Y. A modified evidential methodology of identifying influential nodes in weighted networks// *Physica A.* – 2013. – N 392. – P. 5490 – 5500.
127. Jendoubi S., Martin A. A reliability-based approach for influence maximization using the evidence theory// 19th International Conference on Big Data Analytics and Knowledge Discovery – DaWaK 2017. – Aug 2017, Lyon, France, - 2017. – P.1-14.
128. Zhou K., Pan Q., Martin A. Evidential Community Detection Based on Density Peaks// S. Destercke et al. (Eds.): *BELIEF 2018*, Springer, 2018, LNAI 11069, P .269 – 277.
129. Yager R. On the Dempster-Shafer Framework and New Combination Rules// *Information Sciences.* – 1987. – N 41. – P. 93 – 137.
130. Sentz K., Ferson S. Combination of evidence in Dempster-Shafer theory// Report SAND 2002-0835, Sandia National Laboratories. - 2002. – P. 1-96
131. Newman M.E.J. *Networks: An Introduction.* – Oxford, UK: Oxford University Press.- 2010. –P. 1-720
132. Bonacich P. Technique for Analyzing Overlapping Memberships// *Sociological Methodology.* – 1972. – N 4. – P.176 – 185.
133. Ren J., Wang C., He H., Dong J. Identifying influential nodes in weighted network based on evidence theory and local structure// *Intern. J. of Innovative Comp., Inform. & Control.* – 2015. – N 11(5). – P. 1765 – 1777.
134. Smets P., Kennes R. The Transferable Belief Model// *Artificial Intelligence.* – 1994. – N 66. – P. 191 – 243.
135. Shapley L. A value for n-person games// *Contributions to the Theory of Games. II* (28) in *Annals of Mathematics Studies*, Princeton University Press. – 1953. – P. 307 – 317.
136. Lepskiy A., Meshcheryakova N. Belief Functions for the Importance Assessment in Multiplex Networks// In: Lesot MJ. et al. (eds) *Information Processing and Management of Uncertainty in Knowledge-Based Systems. IPMU 2020.*

Communications in Computer and Information Science. – 2020. – N 1238. – Springer, Cham. – P. 283 – 296.

137. Blei D. Probabilistic topic models// Communications of the ACM. – N 55(4). – 2012. – P. 77 – 84.

138. Song Y., Liu S., Wang H., Wang Z., Li H. Automatic taxonomy construction from keywords// U.S. Patent No. 9,501,569. Washington, DC: U.S. Patent and Trademark Office. - 2016. – P.1-14.

139. Wang C., He X., Zhou A. A Short Survey on Taxonomy Learning from Text Corpora: Issues, Resources and Recent Advances // Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. – 2017. – P. 1190 – 1203.

140. Vedula N., Nicholson P.K., Ajwani D., Dutta S., Sala A., Parthasarathy S. Enriching Taxonomies With Functional Domain Knowledge // The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, ACM. – 2018. P. 745 – 754.

141. Lloret E., Boldrini E., Vodolazova T., Martanez-Barco P., Munoz R., Palomar M. A novel concept-level approach for ultra-concise opinion summarization // Expert Systems with Applications. – 2015. – N 42(20). – P. 7148 – 7156.

142. Mueller G, Bergmann R. Generalization of Workflows in Process-Oriented Case-Based Reasoning // FLAIRS Conference. – 2015. – P. 391 – 396.

143. Waitelonis J., Exeler C., Sack H. Linked data enabled generalized vector space model to improve document retrieval// Proceedings of NLP & DBpedia 2015 workshop in conjunction with 14th International Semantic Web Conference (ISWC), CEUR-WS, N 1486. - 2015. –P.1-15

144. The 2012 ACM Computing Classification System. URL: <http://www.acm.org/about/class/2012> (дата обращения 2018-04-30)

145. Mirkin B., Orlov M. Three aspects of the research impact by a scientist: measurement methods and an empirical evaluation // Optimization, Control, and Applications in the Information Age. – Springer, Cham. – 2015. – P. 233 – 259.

146. Frolov D., Mirkin B., Nascimento S., Fenner T. Finding an appropriate generalization for a fuzzy thematic set in taxonomy// Working paper WP7/2019/01, Moscow, Higher School of Economics Publ. House. – 2018. – 60 p.
147. Nascimento S., Fenner T., Mirkin B. Representing research activities in a hierarchical ontology // Procs. of 3rd International Workshop on Combinations of Intelligent Methods and Applications (CIMA 2012), Montpellier, France, August, 28. – 2012. – P. 23 – 29.
148. Gene ontology consortium: going forward // Nucleic Acids Research, Oxford University Press vol. 43. - 2015. –P.1049-1056.
149. Lee D., Cornet R., Lau F., De Keizer N. A survey of SNOMED CT implementations // Journal of Biomedical Informatics. – 2013. – N 46(1). – P. 87 – 96.
150. Chernyak E. An approach to the problem of annotation of research publications // Proceedings of the eighth ACM international conference on web search and data mining, ACM. – 2015. – P. 429 – 434.
151. Mirkin B., Nascimento S. Additive spectral method for fuzzy cluster analysis of similarity data including community structure and affinity matrices // Information Sciences. – 2012. – N 183(1). – P. 16 – 34.
152. Frolov D., Nascimento S., Fenner T., Mirkin B. Parsimonious Generalization of Fuzzy Thematic Sets in Taxonomies Applied to the Analysis of Tendencies of Research in Data Science // Information Sciences, – 2020. –N 512. – P. 595 – 615.
153. Frolov D., Mirkin B., Nascimento S., Fenner T. Method for Generalization of Fuzzy Sets// Proceedings of the 18th International Conference on Artificial Intelligence and Soft Computing (ICAISC-2019), Springer Lecture Notes in Artificial Intelligence (LNAI). – 2019. – P. 273 – 286.
154. Frolov D., Mirkin B., Nascimento S., Fenner T. Using Taxonomy Tree to Generalize a Fuzzy Thematic Clusters // IEEE 2019 International Conference on Fuzzy Systems Proceedings. IEEE. – 2019.

155. Frolov D., Nascimento S., Fenner T., Taran Z., Mirkin B. Computational Generalization in Taxonomies Applied to: (1) Analyze Tendencies of Research and (2) Extend User Audiences // IDEAL 2019 Proceedings. Springer. – 2020. – P. 3-11.
156. Frolov D., Mirkin B., Nascimento S., Fenner T. Using Domain Taxonomy to Model Generalization of Thematic Fuzzy Clusters// IARIA Content 2019 Proceedings. IARIA. – 2019. – P. 20 – 25.
157. Frolov D., Mirkin B., Nascimento S., Fenner T. Globally Optimal Parsimoniously Lifting a Fuzzy Query Set Over a Taxonomy Tree // Proceedings of the World Congress on Global Optimization, Springer, LNCS. – 2019. – P. 781 – 789.
158. Mirkin B., Frolov D., Vlasov A., Nascimento S., Fenner T. A Hybrid Approach to Interpretable Analysis of Research Paper Collections // Proceedings of the 10th International Conference on Web Intelligence, Mining and Semantics 2020 Jun 30, ACM. – 2020. – P. 184 – 189.
159. Mirkin B., Frolov D., Vlasov A., Nascimento S., Fenner T. A Hybrid Approach to the Analysis of a Collection of Research Papers. In International Conference on Intelligent Data Engineering and Automated Learning 2020 Nov 4, Springer, LNCS. – 2020. – P. 423 – 433.
160. Rubchinsky A. Divisive-Agglomerative Classification Algorithm Based on the Minimax Modification of Frequency Approach// Working paper WP7/2010/07. – M.: NRU HSE. – 2010.
161. Rubchinsky A. Divisive-Agglomerative Algorithm and Complexity of Automatic Classification Problems// Working paper WP7/2015/09. – M.: NRU HSE. – 2015.
162. Rubchinsky A. Family of Graph Decompositions and Its Applications to Data Analysis// Working paper WP7/2016/09. – M.: NRU HSE. – 2016.

163. Rubchinsky A. A New Approach to Network Decomposition Problems// In: Kalyagin V., Nikolaev A., Pardalos P., Prokopyev O. (eds) Models, Algorithms, and Technologies for Network Analysis. NET 2016. Springer Proceedings in Mathematics & Statistics. – N 197. – Springer, Cham. – 2017. –P. 127-152.
164. 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. – N 247. – Springer, Cham. – 2017.-P. 75-111
165. Rubchinsky A. Notion of Graph Entanglement and its Application for Stock Market Analysis// In: Proceedings of 21st IEEE Conference on Business Informatics, Moscow, Russia, 15-17 July 2019. – P.41 – 46
166. Mirkin B.G. Core Concepts in Data Analysis: Summarization, Correlation, Visualization. – Cham, Switzerland: Springer. - 2019.
167. Boginski V., Butenko S., Pardalos P.M. Mining market data: A network approach// Computers & Operations Research. – 2006. – P. 3171 – 3184.
168. Kjær M., Hansen O.H., Thomsen J.P.F. Conceptualizing state capacity// Democracy, the State, and Administrative Reforms Research Report. – 2002. – N 6.
169. Charron N. Diverging Cohesion? Globalisation, State Capacity and Regional Inequalities Within and Across European Countries// European Urban and Regional Studies. – 2016. – N 23(3). – P. 355 – 373.
170. Ахременко А.С., Горельский И.Е., Мельвиль А.Ю. Как и зачем измерять и сравнивать государственную состоятельность разных стран мира? Теоретико-методологические основания// Polis: Journal of Political Studies. – 2019. – N 2. –P. 49-68.
171. Миркин Б.Г. Методы кластер-анализа для поддержки принятия решений: обзор. – 2011.
172. Мячин А.Л. Определение центроидов для повышения точности порядково-инвариантной паттерн-кластеризации// Управление большими системами. – 2019. – N 78. – P. 6 – 22.

173. Ахременко А.С., Мячин А.Л. Паттерн-анализ и кластеризация в исследовании государственной состоятельности: «Адаптивная оптика» для политической науки// Политическая наука. – 2019. – N 3. –P. 112-139.
174. Aleskerov F., Nurmi H. A method for finding patterns of party support and electoral change: An analysis of British general and Finnish municipal elections// Mathematical and Computer Modelling. – 2008. – N 48(9-10). – P. 1385 – 1395.
175. Myachin A.L. Pattern analysis in parallel coordinates based on pairwise comparison of parameters// Automation and Remote Control. – 2019. – N 80. – P. 112 – 123.
176. Inselberg A. Parallel coordinates: visual multidimensional geometry and its applications. – Springer Science & Business Media Vol. 20. - 2009. – P.554.
177. Jain A.K., Murty M.N., Flynn P.J. Data clustering: a review// ACM computing surveys (CSUR). – 1999. – N 31. – P. 264 – 323.
178. Xu R., Wunsch D.C. Survey of clustering algorithms.- 2005.- P.645-678.
179. Dubes R., Jain A.K. Clustering techniques: the user's dilemma// Pattern Recognition. – 1976. – N 8. – P. 247 – 260.
180. Almetova Z., Shepelev V., Shepelev S. Optimization of Delivery Lot Volumes in Terminal Complexes// Transportation Research Procedia. – 2017. – N 27. – P. 396 – 403.
181. Belenky A. Two Classes of Games on Polyhedral Sets in Systems Economics Studies// Network Models in Economics and Finance. – 2014. – Springer. – P. 35 – 84.
182. Bandara Y.M., Nguyen H.-O. Influential factors in port infrastructure tariff formulation, implementation and revision// Transportation Research Part A. – 2016. – N 85. – P. 220 – 232.
183. Kulak O., Genz A., Taner M.E. A decision making tool considering risk assessment of sub-contracting agents for an air cargo shipment planning problem// Journal of Air Transport Management. – 2018. – N 69. – P. 123 – 136.



184. Azadian F., Murat A. Service location grouping and pricing in transportation: Application in air cargo// *European Journal of Operational Research*. 2018. – N 267(3). – P. 933 – 943.
185. Oktal H., Ozger A. Hub location in air cargo transportation: A case study// *Journal of Air Transport Management*. – 2013. – N 27. – P. 1 – 4.
186. Mokhtar H., Redi A.A.N.P., Krishnamoorthy M., Ernst A.T. An inter-modal hub location problem for container distribution in Indonesia// *Computers & Operations Research*. – 2019a. – N 104. – P. 415 – 432.
187. Mokhtar H., Krishnamoorthy M., Ernst A.T. The 2-allocation p-hub median problem and a modified Benders decomposition method for solving hub location problems// *Computers & Operations Research*. – 2019b. – N 104. – P. 375 – 393.
188. Rostami B., Kmmerling N., Buchheim C., Clausen U. Reliable single allocation hub location problem under hub breakdowns// *Computers & Operations Research*. – 2018. – N 96. – P. 15 – 29.
189. Belenky A. Minimax planning problems with linear constraints and methods of their solutions// *Automation and Remote Control*. – 1981. – N 42(10). – P. 1409 – 1419.
190. Belenky A. A planning problem in a class of linear models// *Automation and Remote Control*. – 1978. – N 39(11). – P. 1667 – 1673.
191. Belenky A., Fedin G., Kornhauser A. Optimization models for estimating the volume of investment needed for developing regional infrastructures// XX April International Academic Conference On Economic and Social Development, Moscow, Russia. – 2019.
192. Belenky A., Kornhauser A. Optimally allocating a chain of charging stations for electric vehicles along a highway// 30th European Conference On Operational Research, Dublin, Ireland. – 2019.
193. Hotelling H. Stability in competition// *The Economic Journal*. – 1929. – N 39. – P. 41 – 57.

194. Hakimi S.L. On locating new facilities in a competitive environment// European Journal of Operational Research. – 1983. – N 12. – P. 29 – 35.
195. Plastria F. Static competitive facility location: An overview of optimization approaches// European Journal of Operational Research. – 2001. – N 129(3). – P. 461 – 470.
196. 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.
197. Kress D., Pesch E. Sequential competitive location on networks// European Journal of Operational Research. – 2012. – N 217(3). – P. 483 – 499.
198. 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.
199. Küçükaydin H., Aras N., Kuban Altinel I. Competitive facility location problem with attractiveness adjustment of the follower: A bi-level programming model and its solution// European Journal of Operational Research. – 2011. – N 208(3). – P. 206 – 220.
200. 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.
201. 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.
202. Fernandez J., Salhi S., Toth B.G. Location equilibria for a continuous competitive facility location problem under delivered pricing// Computers & Operations Research. – 2014. – N 41. – P. 185 – 195.
203. Zhang Y. Designing a retail store network with strategic pricing in a competitive environment// Int. J. Production Economics. – 2015. – N 159. – P. 265 – 273

204. Adjiman C.S., Androulakis I.P., Floudas C.A. Global optimization of MINLP in process synthesis and design// Computers and Chemical Engineering. – 1997. – N 21. – P. 445 – 450.
205. Fleckinger P., Lafay T. Product flexibility and price competition in Hotelling's duopoly// Mathematical Social Sciences. – 2010. – N 60. – P. 61 – 68.
206. 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.
207. Dasci A., Laporte G. A continuous model for multistore competitive location// Operations Research. – 2005. – N 53. – P. 263 – 280.
208. Dasci A., Laporte G. Erratum to "A continuous model for multistore competitive location"// Operations Research. – 2007. – N 55, – P. 1188 – 1190.
209. Marianov V., Serra D., ReVelle C. Location of hubs in a competitive environment// European Journal of Operational Research. – 1999. – N 114. – P. 363 – 371.
210. Wagner B. A note on "Location of hubs in a competitive environment"// European Journal of Operational Research. – 2008. – N 184. – P. 57 – 62.
211. Eiselt H.A., Marianov V. A conditional p-hub location problem with attraction functions// Computers & Operations Research. – 2009. – N 36. – P. 3128 – 3135.
212. 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.
213. Mahmutogullari A.I., Kara B.Y. Hub location under competition// European Journal of Operational Research. – 2016. – N 250. – P. 214 – 225.
214. Luer-Villagra A., Marianov V. A competitive hub location and pricing problem// European Journal of Operational Research. – 2013. – N 231. – P. 734 – 744.

215. Belenky A., Kornhauser A. Estimating the needed volume of investment in developing a chain of charging stations for electric vehicles along a highway// *International Journal of Public Administration*. – 2019b. – N 42. – P. 1256 – 1274.
216. Belenky A., Larson R., Roginskiy L. Storing electricity in a country's electrical grid as a key energy problem of the 21st century// *Procedia Computer Science*. – 2016. – N 91. – P. 873 – 881.
217. Belenky A. A game-theoretic approach to optimizing the scale of incorporating renewable sources of energy and electricity storing systems in a regional electrical grid// *Energy Systems*. – 2015. – N 6(3). – P. 389 – 415.
218. Zhang Y., Zhang Q., Farnoosh A., Chen S., Li Y. CIS-based multi-objective particle swarm optimization of charging stations for electric vehicles// *Energy*. – 2019. – N 169. – P. 844 – 853.
219. Jamian J., Mustafa M., Mokhlis H., Baharudin M. Simulation study on optimal placement and sizing of battery switching station units using Artificial Bee Colony algorithm// *Electrical Power and Energy Systems*. – 2014. – N 55. – P. 592 – 601.
220. Wang Y., Shi J., Wang R., Liu Z., Wang L. Siting and sizing of fast charging stations in highway network with budget constraint// *Applied Energy*. – 2018. – N 228. – P. 1255 – 1271.
221. He Y., Kockelman K., Perrine K. Optimal locations of U.S. fast charging stations for long-distance trip completion by battery electric vehicles// *Journal of Cleaner Production*. – 2019. – N 214. – P. 452 – 461.
222. Erbas M., Kabak M., Ozceylan E., Cetinskaya C. Optimal siting of electric vehicles charging stations: A CIS-based fuzzy Multi-Criteria Decision Analysis// *Energy*. – 2018. – N 163. – P. 1017 – 1031.
223. Levinson R., West T.H. Impact of public electric vehicle charging infrastructure// *Transportation Research Part D: Transport and Environment*. – 2018. – N 64. – P. 158 – 177.

224. Wang C., Dunn R., Robinson F., Lian B., Yuan W., Redfern M. Active-reactive power approaches for optimal placement of charge stations in power systems// *Electrical Power and Energy Systems*. – 2017. – N 84. – P. 87 – 98.
225. Gnann T., Funke S., Jakobsson N., Plotz P., Bennehag A. Fast charging infrastructure for electric vehicles: Today's situation and future needs// *Transportation Research Part D: Transport and Environment*. – 2018. – N 62. – P. 314 – 329.
226. Arslan O., Karasan O. A Benders decomposition approach for the charging station location problem with plug-in hybrid electric vehicles// *Transportation Research Part B: Methodological*. – 2016. – N 93. – P. 70 – 695.
227. Shi R., Lee K. Multi-objective optimization of electric vehicle fast charging stations with SPEA-II// *In IFAC-Papers-Online*. – 2015. – N 48(30). – P. 535 – 540.
228. Li S., Huang Y., Mason, S. A multi-period optimization model for the development of public electric vehicle charging stations on network// *Transportation Research Part C: Emerging Technologies*. – 2016. – N 65. – P. 128 – 143.
229. Brooker R., Qin N. Identification of potential locations of electric vehicle supply equipment// *Journal of Power Sources*. – 2015. – N 299. – P. 76 – 84.
230. Xie F., Liu C., Li S., Lin Z., Huang Y. Long-term strategic planning of inter-city charging infrastructure for battery electric vehicles// *Transportation Research Part E: Logistics and Transportation Review*. – 2018. – N 109. – P. 261 – 276.
231. Tao Y., Huang M., Yang L. Data-driven optimized layout of battery electric vehicle charging infrastructure// *Energy*. – 2018. – N 150. – P. 735 – 744.
232. Asamer J., Reinthaler M., Ruthmair M., Straub M., Puchinger J. Optimizing charging station locations for urban taxi providers// *Transportation Research Part A: Policy and Practice*. – 2016. – N 85. – P. 233 – 246.
233. Funke S., Plotz P., Wietschel M. Invest in fast-charging infrastructure or in longer battery ranges? A cost-efficiency comparison for Germany// *Applied Energy*. – 2019. – N 235. – P. 888 – 899.

234. Guo Z., Deride J., Fan Y. Infrastructure planning for fast charging stations in a competitive market// *Transportation Research Part C*. – 2016. – N 68. – P. 215 – 227.
235. Rao R., Zhang X., Xie J., Ju L. Optimizing electric vehicle users' charging behavior in battery swapping mode// *Applied Energy*. – 2015. – N 155. – P. 547 – 559.
236. Yang J., Sun H. Battery swap station location-routing problem with capacitated electric vehicles// *Computers and Operations Research*. – 2015. – N 55. – P. 217 – 232.
237. Adler J., Mirchandani P. Online routing and battery reservations for electric vehicles with swappable batteries// *Transportation Research Part B: Methodological*. – 2014. – N 70. – P. 285 – 302.
238. Jing W., Kim I., An K. The Uncapacitated Battery Swapping Facility Location Problem with Localized Charging System Serving Electric Bus Fleet// *Transportation Research Procedia*. – 2018. – N 34. – P. 227 – 234.
239. Yang J., Dong J., Hu L. A data-driven optimization-based approach for siting and sizing of electric taxi charging stations// *Transportation Research Part C*. – 2017. – N 77. – P. 462 – 477.
240. Amiri S., Jadid S., Saboori H. Multi-objective optimum charging management of electric vehicles through battery swapping stations// *Energy*. – 2018. – N 165(B). – P. 549 – 562.
241. Shahraki N., Cai H., Turkay M., Xu M. Optimal locations of electric public charging stations using real world vehicle travel patterns// *Transportation Research Part D*. – 2015. – N 41. – P. 165 – 176.
242. FRAM 2018. The website of FRAM annual publication. – 2018. URL: <https://issuu.com/framcentre/docs/framforum-2018-issuu> (дата обращения 2020-05-25)
243. Wilson G.N. Abandoning the nest: regional mergers and their impact on the Russian North// *Polar Geography*. – 2003. – 27(3). – P. 240 – 253.

244. Arctic Council. Arctic Council Secretariat annual report 2018. Arctic Council Secretariat. 2019. [Electronic resource]. 2019. URL: <https://oaarchive.arctic-council.org/handle/11374/2340> (02.06.20)

245. Global Migration Data Portal 2020. The website of migration data portal URL: [https://migrationdataportal.org/?t=2019&i=stock\\_abs\\_&cm49=643](https://migrationdataportal.org/?t=2019&i=stock_abs_&cm49=643) (дата обращения 2020-05-10)

246. Heleniak T. From Migrants to Workers: International migration trends in the Nordic countries. – Stockholm: Nordregio. – 2018.. URL: <http://norden.diva-portal.org/smash/record.jsf?pid=diva2%3A1186222&dswid=8412> (дата обращения 2020-05-10)

247. AHDR. New Report: Arctic Human Development Report – volume II. // The website UArctic. URL: <https://www.uarctic.org/news/2015/2/new-report-arctic-human-development-report-volume-ii-published/> (дата обращения 2020-05-20)

248. UN Population Fund. UNFPA Global Situation Report No1. 2020. The website of UNFPA. URL: <https://www.unfpa.org/resources/global-covid-19-situation-report-no-1> (дата обращения 2020-06-20)

249. Heleniak T.W. Cities on Ice: Population change in the Arctic// Norgregio. – 2018.

250. Nordregio Polar Peoples in the Future: Projections of the Arctic Population executive summary. The website of Nordregio – 2020. URL: <https://nordregio.org/publications/polar-peoples-projections-of-the-arctic-population-executive-summary/> (дата обращения 2020-06-02)

251. The Arctic Council: A backgrounder, Arctic People.]. 2020. URL: <https://arctic-council.org> (дата обращения 2020-05-10)

252. Donald L., Gautier et. al. Assessment of Undiscovered Oil and Gas in the Arctic// Science. – 2009. – N 324. – P. 1175 – 1179.

253. Bird K.J. et al. Circum-Arctic resource appraisal: Estimates of undiscovered oil and gas north of the Arctic Circle// Fact sheet. – 2008. – N 3049. – P.1-

254. Conservation of Arctic Flora and Fauna (CAFF). Arctic Flora and Fauna: Status and Conservation. – Edita, Helsinki. – 2001. – 266 pp.
255. Shvydun S., Aleskerov F. A mathematical approach to conflict resolution in the Arctic Region// In: 2017 3rd IEEE international conference on cybernetics (CYBCONF), Exeter. – 2017. – P. 1 – 6.
256. Shvydun S., Aleskerov F. Allocation of Disputable Zones in the Arctic Region// Group Decision and Negotiation. – 2019. – N 28. – P. 11 – 42.
257. Aleskerov F., Victorova E. An analysis of potential conflict zones in the Arctic Region// Working paper WP7/2015/05. – Moscow: HSE Publishing House. – 2015. – 20 p.
258. Будыко М.И. Некоторые пути воздействия на климат// Метеорология и гидрология. – 1962. – N 2. – С. 3 – 8.
259. Chapin III F.S. et al. Building resilience and adaptation to manage Arctic change// AMBIO: A Journal of the Human Environment. – 2006. – N 35(4). – P. 198 – 202.
260. Hamilton L.C. et al. Climigration? Population and climate change in Arctic Alaska// Population and environment. – 2016. – N 38(2). – P. 115 – 133.
261. Ebinger C.K., Zambetakis E. The geopolitics of Arctic melts// International Affairs. – 2009. – N 85(6). – P. 1215 – 1232.
262. Brubaker M., Chavan P., Berner J, Warren M. Climate Change in Selawik, Alaska// Alaska Native Tribal Health Consortium (ANTHC). – 2012. – 46 p.
263. Пискулова Н.А. Киотский протокол: возможности для России. – М.: МГИМО, WWF России. – 2006. 90 с.
264. Ashton P. Avoiding Conflicts over Africa's Water Resources// Ambio A Journal of the Human Environment. – 2002. – T31(3). – P. 236 – 242.
265. ФАО. Водные ресурсы и обеспечение продовольственной безопасности и питания// Доклад Группы экспертов высокого уровня по вопросам



продовольственной безопасности и питания Комитета по всемирной продовольственной безопасности. – 2015. – Рим. – 129 с. URL: <http://www.fao.org/3/a-av045r.pdf> (дата обращения 2020-05-10)

266. JMP, Progress on Drinking Water and Sanitation: Special Focus on Sanitation// World Health Organization and United Nations Children's Fund Joint Monitoring Programme for Water Supply and Sanitation. – Geneva and New York. – 2008. P. 1-138.

267. WHO & UNICEF. Meeting the MDG drinking water and sanitation target: The urban and rural challenge of the decade', World Health Organization, Geneva. – 2006. – 41 p.

268. Smiley S.L. Defining and measuring water access: lessons from Tanzania for moving forward in the post-Millennium Development Goal era// African Geographical Review. – 2017. – 36(2). – P. 168 – 182.

269. Xu H., Wu M. Water Availability Indices – A Literature Review// Energy Systems Division, Argonne National Laboratory. – 2017. – 44 p.

270. CIA 2020, The world factbook – Algeria, CIA, retrieved 20 July 2020, <<https://www.cia.gov/library/publications/the-world-factbook/geos/ag.html>>

271. ISI Water n.d., Water Scarcity: A growing concern across the globe// ISI Water.- URL:<https://isi-water.com/solving-water-scarcity/#:~:text=As%20the%20largest%20country%20in,even%20then%20it%20is%20minimal.&text=The%20bulk%20of%20Algeria's%20water,the%20result%20of%20unprecedented%20drought.>(дата обращения 2020-07-20)

272. Sleet P. Water Protests in Algeria Are Giving Cause for Concern about its Long-Term Stability// Future Directions International Pty Ltd. 2019. URL:<https://www.futuredirections.org.au/publication/water-protests-in-algeria-are-giving-cause-for-concern-about-its-long-term-stability/>( дата обращения 2020-07-20)

273. Benabdallah S. The water resources and water management regimes in Tunisia, Agricultural Water Management// Proceedings of a Workshop in Tunisia (Series: Strengthening Science-Based Decision Making in Developing Countries). – The National Academies Press, Washington, DC. – 2007. – P. 81 – 87
274. Paca J. et al. Quality assessment of water intended for human consumption from Kwanza, Dande and Bengo rivers (Angola)// Environmental Pollution. – 2019. – N 254(B). – 113037.
275. Cain A., Baptista A. Community Management and the Demand for ‘Water for All’ in Angola’s Musseques// Water. – 2020. – 12(6). – P. 1 – 19.
276. Fanack water 2019, Water quality and Use in Morocco// Fanack water. – 2019. URL:<https://water.fanack.com/morocco/water-quality-and-use-morocco/>(дата обращения 2020-04-11)
277. UNEP. Africa Water Atlas. Division of Early Warning and Assessment (DEWA)// United Nations Environment Programme (UNEP), Nairobi, Kenya. – 2010.URL:[https://na.unep.net/atlas/africaWater/downloads/africa\\_water\\_atlas.pdf](https://na.unep.net/atlas/africaWater/downloads/africa_water_atlas.pdf) (дата обращения 2019-12-01)
278. Karekezi S., Kithyoma W. (ed.). Sustainable Energy in Africa: Cogeneration and Geothermal in East and Horn of Africa: Status and Prospects. – AFREPREN/FWD. - 2005. – P.1-30.
279. Thornton P.K., van de Steeg J., Notenbaert A., Herrero M. The impacts of climate change on livestock and livestock systems in developing countries: A review of what we know and what we need to know// Agricultural systems. – 2009. – N 101(3). – P. 113 – 127.
280. Urama K.C., Ozor N. Impacts of climate change on water resources in Africa: the role of adaptation// African Technology Policy Studies Network. – 2010. – N 29. – P. 1 – 29.
281. World Meteorological Organization. State of the Climate in Africa// WMO. – 2020. – N 1253. – P. 1 – 37.
282. Hulme M., Doherty R., Ngara T., New M., Lister D. African climate change: 1900-2100// Climate research. – 2001. – 17(2). – P. 145 – 168.

283. Collier P., Conway G., Venables T. Climate change and Africa// Oxford Review of Economic Policy. – 2008. – N 24(2). – P. 337 – 353.
284. Funk C., Senay G., Asfaw A., Verdin J., Rowland J., Michaelson J., Choularton R. Recent drought tendencies in Ethiopia and equatorial-subtropical eastern Africa// Washington DC, FEWS-NET. – 2005.-P. 1011-1027.
285. Schlenker W., Lobell D.B. Robust negative impacts of climate change on African agriculture// Environmental Research Letters. – 2010. – N 5(1). – P. 014010.
286. Thornton P.K., Jones P.G., Ericksen P.J., Challinor A.J. Agriculture and food systems in sub-Saharan Africa in a 4 C+ world// Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. – 2011. – N 369(1934). – P. 117 – 136.
287. Lloyd S.J., Kovats R.S., Chalabi Z. Climate change, crop yields, and undernutrition: development of a model to quantify the impact of climate scenarios on child undernutrition// Environmental health perspectives. – 2011. – N 119(12). – P. 1817 – 1823.
288. Guha-Sapir D., Hoyols P. Measuring the human and economic impact of disasters// CRED, Government Office for Science. – 2012.-P. 1-50.
289. Huang J., Liu Y., Ma L. Assessment of regional vulnerability to natural hazards in China using a DEA model// Int. Journal Disaster Risk Science. – 2011. – 2(2). – P. 41 – 48.
290. Charnes A., Cooper W. Programming with linear fractional functional// Naval Research Logistics Quarterly. – 1962. – N 9. – P. 181–186.
291. Charnes A., Cooper W.W., Rhodes E. Measuring the Efficiency of Decision Making Units// European Journal of Operational Research. – 1978. – N 2. – P. 429 – 444.
292. Yucel L. Measuring the Efficiency of Portfolios with Data Envelopment Analysis// Sosyal Bilimler Dergisi. – 2010. – N 2. – P. 116 – 121.
293. Sanchez L., Vasquez C., Vilorio A. The Data Envelopment Analysis to Determine Efficiency of Latin American Countries for Greenhouse Gases Control

in Electric Power Generation// International Journal of Energy Economics and Policy. – 2018. – N 8(3). – P. 197 – 208.

294. Smriti T.N., Khan Md H.R. Efficiency Analysis of Manufacturing Firms Using Data Envelopment Analysis Technique// Journal of Data Science. – 2018. – N 16(1). – P. 69 – 78.

295. Li C.-H., Li N., Wu L.-C., Hu A.-J. A relative vulnerability estimation of flood disaster using data envelopment analysis in the Dongting Lake region of Hu-nan// Natural hazards and earth system sciences. – 2013. – N 13(7). – P. 1723 – 1734

296. Aleskerov F., Demin S. An Assessment of the Impact of Natural and Technological Disasters Using a DEA Approach, Dynamics of Disasters – Key Concepts, Models, Algorithms, and Insights. – Switzerland: Springer. – 2016. P. 1 – 14.

297. Cheng H.-T., Chang H.-S. A Spatial DEA-Based Framework for Analyzing the Effectiveness of Disaster Risk Reduction Policy Implementation: A Case Study of Earthquake-Oriented Urban Renewal Policy in Yongkang, Taiwan// Sustainability. – 2018. – N 10. – P. 1751.

298. Yu X., Chen H., Li C. Evaluate Typhoon Disasters in 21st Century Maritime Silk Road by Super-Efficiency DEA// International Journal of Environmental Research and Public Health. – 2019. – N 16(9). – P. 1614.

299. De Almada Garcia Adriano P., Curty Leal Junior I., Alvarenga Oliveira M. A weight restricted DEA model for FMEA risk prioritization// Producao. – 2013. – N 23(3). – P. 500 – 507.

300. Aleskerov F. Multicriterial Interval Choice Models// Information Sciences. – 1994. – N 80. – P. 25 – 41.

301. Aleskerov F., Demin S. DEA for the Assessment of Regions Ability to Cope with Disasters// in: Dynamics of Disasters 2019, Springer. – 2020. (в печати)

302. Борисова О.В., Ворожцова В.В. Теневой банкинг и целесообразность его регулирования в России// Аудит и финансовый анализ. – 2015. – N 2. – С. 218 – 223.

303. FSB. Global Shadow Banking Monitoring Report 2017// 2018. URL:<https://www.fsb.org/2018/03/global-shadow-banking-monitoring-report-2017/> (дата обращения 2020-05-10).
304. Ведомости. Дыра в капитале группы «Открытие» составила 300 млрд рублей// Ведомости, 30 ноября 2017 URL: <https://www.vedomosti.ru/finance/articles/2017/11/30/743700-dira-v-kapitale-gruppi-otkritiya-300-mlrd> (дата обращения 2020-07-20)
305. Ordoñez G. Sustainable Shadow Banking// American Economic Journal: Macroeconomics. – 2018. – N 10(1). – P. 33 – 56.
306. Buchak G., Matvos G., Piskorski T., Seru A. Beyond the balance sheet model of banking: implications for bank regulation and monetary policy// National Bureau of Economic Research (NBER). – 2020. –P.18-75
307. Пеникас Г.И. Проектирование оптимальной системы финансового регулирования рисков: опыт регулирования дорожного движения (Часть 1)// Управление финансовыми рисками. – 2016. – 48(4). – P. 242 – 254.
308. Mishin A. Dynamic Bank Capital Regulation and Optimal Macroprudential Policies in the Presence of Shadow Banks// Job Market Paper. – 2020. – January. [https://c722622c-dda4-4fcf-9af5-c82f7cb1076a.filesusr.com/ugd/f22857\\_3c3dad4e108a46d899c75b93130b3c43.pdf](https://c722622c-dda4-4fcf-9af5-c82f7cb1076a.filesusr.com/ugd/f22857_3c3dad4e108a46d899c75b93130b3c43.pdf).
309. 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.
310. Aleskerov F.T., Makarenkov V., Tverskoi D. Modeling functional specialization of a cell colony under different fecundity and viability rates and resource constraint// PloS one – 2018. – N 2018-08-08. P. 16.
311. Dummit D.S., Foote R.M. Abstract Algebra. – 2nd ed., Englewood Cliffs, NJ: Prentice- Hall. – 2003.
312. Elizarova T.G. Quasi-Gas Dynamic Equations. – Springer: Dordrecht. – 2009.-P. 37-62.

## ПРИЛОЖЕНИЕ А

### Таблицы к подразделу 1.5

Таблица А.1 – Вопросы анкет, использовавшиеся для исследования

Вопрос (шкала ответов)	Шкала ответов	Варианты ответов
Укажите Ваш курс обучения	Число от 1 до 9	<ol style="list-style-type: none"> <li>1 курс бакалавриата/специалитета</li> <li>2 курс бакалавриата/специалитета</li> <li>3 курс бакалавриата/специалитета</li> <li>4 курс бакалавриата</li> <li>4 курс специалитета</li> <li>5 курс специалитета</li> <li>6 курс специалитета</li> <li>1 курс магистратуры</li> <li>2 курс магистратуры</li> </ol>
Оцените, насколько Вы удовлетворены следующими аспектами обучения в НИУ ВШЭ	Для каждого пункта шкала от 1 до 5, где «1» — абсолютно не удовлетворен(а), «5» — полностью удовлетворен(а), «99» - затрудняюсь ответить	<ol style="list-style-type: none"> <li>1. Качество получаемого образования (обязательная программа)</li> <li>2. Качество научно-образовательных мероприятий, реализуемых вне рамок обязательной программы (открытые лекции, мастер-классы, конференции, семинары и т.д.)</li> <li>3. Полезность получаемого образования для будущей работы / профессиональной деятельности</li> <li>4. Уровень преподавания английского языка в Вышке</li> <li>1. Условия для аудиторного обучения (оснащенность, вместимость аудиторий и т.д.);</li> <li>2. Условия для самостоятельных занятий (доступность и оснащенность компьютерных классов, аудиторий для самостоятельных занятий, коворкингов и т.д.);</li> <li>3. Информационное сопровождение обучения (информирование о расписании занятий, экзаменов и других организационных вопросов);</li> <li>4. Доступность учебной литературы и материалов для обучения (включая ридеры, учебники из обязательной программы)</li> </ol>
Если бы Вы могли вновь принимать решение о поступлении в вуз, то что бы Вы выбрали?	Для каждого пункта число 0 или 1	<ol style="list-style-type: none"> <li>1. Учиться в НИУ ВШЭ на той же образовательной программе, на которой учусь сейчас</li> <li>2. Учиться в НИУ ВШЭ, но на другой образовательной программе</li> <li>3. Не поступать в НИУ ВШЭ, а выбрать другой вуз (укажите, какой именно)</li> <li>4. Вообще не поступать в вуз</li> <li>5. Затрудняюсь ответить</li> </ol>
Скажите, пожалуйста, включали ли Вы в прошлом или текущем учебном году следующие виды дисциплин в свой учебный план?	Для каждого пункта число 0 или 1	<ol style="list-style-type: none"> <li>1. Онлайн-дисциплины (все занятия и экзамен проходили в онлайн формате)</li> <li>2. Дисциплины со смешанным обучением, где нужно было прослушивать онлайн-лекции и посещать оффлайн-семинары</li> <li>3. Дисциплины со смешанным обучением, где не было оффлайн-семинаров (очно проходил только экзамен)</li> <li>4. Нет, не включал(а) никакие из перечисленных дисциплин в свой учебный план</li> </ol>

## Продолжение Таблицы А.1

Вопрос (шкала ответов)	Шкала ответов	Варианты ответов
Какой формат учебного курса предпочтительнее для Вас: онлайн-курс или курс, предполагающий аудиторные занятия с преподавателем?	Число от 1 до 6 или 99	<ol style="list-style-type: none"> <li>1. Определенно предпочитаю онлайн-курсы</li> <li>2. Скорее предпочитаю онлайн-курсы</li> <li>3. Одинаково отношусь к обоим форматам</li> <li>4. Предпочитаю курсы, сочетающие в себе онлайн занятия с аудиторными занятиями с преподавателем</li> <li>5. Скорее предпочитаю курсы, предполагающие аудиторные занятия с преподавателем</li> <li>6. Определенно предпочитаю курсы, предполагающие аудиторные занятия с преподавателем</li> <li>99. Затрудняюсь ответить</li> </ol>
Оцените, пожалуйста, степень согласия со следующими утверждениями относительно организации учебного процесса, связанного с онлайн-дисциплинами (в анкете 2019 года предлагалось только выбрать те трудности, с которыми столкнулись студенты)	Для каждого пункта шкала от 1 до 5, где «1» - абсолютно не согласен(-на), «5» - полностью согласен(-на), «99» - затрудняюсь ответить	<ol style="list-style-type: none"> <li>1. Список онлайн-дисциплин, предложенный студентам для освоения, короткий (вариантов недостаточно) и некачественный (нет интересующих меня дисциплин)</li> <li>2. Существуют проблемы с переводом оценок за онлайн-дисциплины в десятибалльную систему оценок НИУ ВШЭ</li> <li>3. Мне пришлось самостоятельно оплатить прохождение онлайн-дисциплины, которую я включал(а) в свой учебный план</li> <li>4. При прохождении экзамена по онлайн-дисциплине студенту легко списать</li> <li>5. Учебный офис образовательной программы допускает ошибки в работе с онлайн-курсами, в том числе предоставляет студентам доступ к онлайн-курсам позже необходимых сроков и т.п.</li> <li>6. Программа онлайн-лекций не соответствовала программе оффлайн-семинаров</li> </ol>
Как Вы оцениваете идею перевести все лекционные занятия в Вышке в онлайн формат? (в анкете 2018 года, в анкете 2019 года не было такого вопроса)	Число от 1 до 5	<ol style="list-style-type: none"> <li>1. Полностью поддерживаю</li> <li>2. Скорее поддерживаю</li> <li>3. Скорее не поддерживаю</li> <li>4. Точно не поддерживаю</li> <li>5. Затрудняюсь ответить</li> </ol>
При прочих равных условиях какой формат учебного курса Вы бы выбрали? (в анкете 2019 года, в анкете 2018 года не было такого вопроса)	Выбор из предложенных ответов	<ol style="list-style-type: none"> <li>1. Онлайн-курсы</li> <li>2. Дисциплины со смешанным обучением</li> <li>3. Оффлайн-курсы</li> <li>99. Мне всё равно<sup>8</sup></li> </ol>

<sup>8</sup> Этот вариант не выбран никем из тех, кому был предложен этот вопрос (то есть из числа тех, кто включал в прошлом или текущем учебном году онлайн-дисциплины или blended-дисциплины в свой учебный план).

## Продолжение Таблицы А.1

Вопрос (шкала ответов)	Шкала ответов	Варианты ответов
По каким причинам Вы не поддерживаете данную идею? (в анкете 2019 года не было такого вопроса)	Для каждого пункта число 0 или 1	<ol style="list-style-type: none"> <li>1. Для меня важно личное общение с преподавателем</li> <li>2. Для меня важно личное общение с другими студентами</li> <li>3. Для меня важно наличие четкого графика занятий</li> <li>4. В аудитории я лучше воспринимаю информацию</li> <li>5. Другое</li> </ol>
Есть ли у Вас оплачиваемая работа?	Для каждого пункта число 0 или 1	<ol style="list-style-type: none"> <li>1. работа на полный день в Вышке</li> <li>2. работа на полный день вне Вышки</li> <li>3. работа на неполный день в Вышке</li> <li>4. работа на неполный день вне Вышки</li> <li>5. нерегулярные приработки в Вышке</li> <li>6. нерегулярные приработки вне Вышки</li> </ol>

Таблица А.2 – Значимые вопросы и ответы по кластерам

№ кластера	Отношение к переводу лекций в онлайн	Вопросы, повлиявшие на предпочтения относительно перевода лекций в онлайн формат	Ответы, повлиявшие на предпочтения относительно перевода лекций в онлайн формат
1	за	Если бы Вы могли вновь принять решение о поступлении в вуз, то, что бы Вы выбрали?	Учиться в НИУ ВШЭ на той же образовательной программе, на которой учусь сейчас
2	за	Какой формат учебного курса предпочтителен для Вас: онлайн-курс или курс, предполагающий аудиторные занятия с преподавателем?	<ol style="list-style-type: none"> <li>1. Определенно предпочитаю онлайн-курсы</li> <li>2. Скорее предпочитаю онлайн-курсы</li> </ol>
3	против	По каким причинам Вы не поддерживаете идею перевести все лекционные занятия в Вышке в онлайн формат?	<ol style="list-style-type: none"> <li>1. Для меня важно наличие четкого графика занятий</li> <li>2. В аудитории я лучше воспринимаю информацию</li> </ol>
4	против	По каким причинам Вы не поддерживаете идею перевести все лекционные занятия в Вышке в онлайн формат?	<ol style="list-style-type: none"> <li>1. Для меня важно личное общение с преподавателем</li> <li>2. Для меня важно личное общение с другими студентами</li> <li>3. В аудитории я лучше воспринимаю информацию</li> </ol>
5	за	Оцените, пожалуйста, степень согласия с утверждением: существуют проблемы с переводом оценок за онлайн-дисциплины в десятибалльную систему оценок НИУ ВШЭ	Абсолютно не согласен(-на)



## Продолжение Таблицы А.2

№ кластера	Отношение к переводу лекций в онлайн	Вопросы, повлиявшие на предпочтения относительно перевода лекций в онлайн формат	Ответы, повлиявшие на предпочтения относительно перевода лекций в онлайн формат
6	скорее против	<ol style="list-style-type: none"> <li>1. Оцените, насколько Вы удовлетворены уровнем преподавания английского языка в Вышке</li> <li>2. Оцените, насколько Вы удовлетворены условиями для самостоятельных занятий</li> <li>3. Если бы Вы могли вновь принимать решение о поступлении в вуз, то, что бы Вы выбрали?</li> <li>4. По каким причинам Вы не поддерживаете идею перевести все лекционные занятия в Вышке в онлайн формат?</li> </ol>	<p>Нет ответов</p> <p>Полностью удовлетворен(а)</p> <p>Учиться в НИУ ВШЭ на той же или на другой образовательной программе, на которой учусь сейчас Учиться в НИУ ВШЭ на той же магистерской программе, на которой учусь сейчас</p> <ol style="list-style-type: none"> <li>1. Для меня важно личное общение с преподавателем</li> <li>2. Для меня важно личное общение с другими студентами</li> <li>3. Для меня важно наличие четкого графика занятий</li> <li>4. В аудитории я лучше воспринимаю информацию</li> </ol>

Таблица А.3 – Описание кластеров, вопросы про онлайн формат

Кластер	Брали ли онлайн курсы	Предпочтительный формат курсов	Отношение к высказываниям против онлайн формата
1	никто не брал онлайн-дисциплины и смешанные (в обоих форматах)	Не предлагался такой вопрос	Не предлагался такой вопрос
2	все брали онлайн, не брали смешанные (онлайн-лекции и оффлайн-семинары), и почти все не брали смешанные (онлайн-лекции без оффлайн-семинаров, очно только экзамен)	Большинство за смешанный формат: 33% за оффлайн-курсы, 40% за дисциплины со смешанным обучением, 14% за онлайн-курсы	Все ответили «я не сталкивался(-ась) ни с какими трудностями»
3	большинство брали онлайн-дисциплины, все брали смешанные (онлайн-лекции и оффлайн-семинары), почти все не брали смешанные (онлайн-лекции без оффлайн-семинаров, очно только экзамен)	Большинство за смешанный формат: 39% за оффлайн-курсы, 45% за дисциплины со смешанным обучением, 6% за онлайн-курсы	Все ответили «я не сталкивался(-ась) ни с какими трудностями»

Продолжение Таблицы А.3

Клас-стер	Брали ли онлайн курсы	Предпочтительный формат курсов	Отношение к высказываниям про-тив онлайн формата
4	все брали онлайн, не брали смешанные (онлайн-лекции и оффлайн-семинары), почти все не брали смешанные (онлайн-лекции без оффлайн-семинаров, очно только экзамен)	За оффлайн и за смешанный формат: 40% за оффлайн-курсы, 40% за дисциплины со смешанным обучением, 13% за онлайн-курсы	Примерно половина отмечает узкий список онлайн-дисциплин; больше, чем в других классах (но на уровне 10%) отмечают «мне пришлось самостоятельно оплатить курс»; больше, чем в других классах (на уровне 30%) отмечают «учебный офис допускал ошибки в работе с онлайн-курсами»; больше, чем в других классах (но на уровне 10%) отмечают «программа дисциплины отсутствовала»; на уровне 30% отмечают «мне было непонятно, у кого можно проконсультироваться»
5	большинство брали онлайн-дисциплины и не брали смешанные (онлайн-лекции и оффлайн-семинары), почти все не брали смешанные (онлайн-лекции без оффлайн-семинаров, очно только экзамен)	Большинство за оффлайн: 46% за оффлайн-курсы, 39% за дисциплины со смешанным обучением, 9% за онлайн-курсы	Больше, чем в других классах (но на уровне 10%) отмечают проблемы с переводом оценок/начислением кредитов; единственные, кто (на уровне 30%) отмечают «узкий список онлайн-дисциплин»; больше, чем в других классах (на уровне 30%) отмечают «учебный офис допускал ошибки в работе с онлайн-курсами»; на уровне 30% отмечают «программа лекций не соответствовали программе семинаров»; больше, чем в других классах (но на уровне 10%) отмечают «программа дисциплины отсутствовала»; на уровне 30% отмечают «мне было непонятно, у кого можно проконсультироваться»; больше, чем в других классах (на уровне 30%) отмечают «преподаватель оффлайн-семинаров не знал содержания лекций»
6	никто не брал онлайн-дисциплины и смешанные (онлайн-лекции и оффлайн-семинары), все брали смешанные (онлайн-лекции без оффлайн-семинаров, очно только экзамен)	За оффлайн и за смешанный формат: 41% за оффлайн-курсы, 39% за дисциплины со смешанным обучением, 8% за онлайн-курсы	на уровне 30% отмечают «узкий список онлайн-дисциплин»; на уровне 30% отмечают «мне было непонятно, у кого можно проконсультироваться»
7	большинство брали онлайн-дисциплины и все брали смешанные (онлайн-лекции и оффлайн-семинары), почти все не брали смешанные (онлайн-лекции без оффлайн-семинаров, очно только экзамен)	Большинство за смешанный формат: 37% за оффлайн-курсы, 47% за дисциплины со смешанным обучением, 12% за онлайн-курсы	Все ответили «я не сталкивался(-ась) ни с какими трудностями»

# Продолжение Таблицы А3

Класс	Брали ли онлайн курсы	Предпочтительный формат курсов	Отношение к высказываниям против онлайн формата
8	никто не брал онлайн-дисциплины и смешанные (в обоих форматах)	Не предлагался такой вопрос	Не предлагался такой вопрос

Таблица А.4 – Результат работы дисперсионного анализа

№ класса	Предпочтительный формат курса	Вопросы, повлиявшие на предпочтения относительно перевода лекций в онлайн формат	Ответы, повлиявшие на предпочтения относительно перевода лекций в онлайн формат
2	Большинство за смешанный формат: 33% за оффлайн-курсы, 40% за дисциплины со смешанным обучением, 14% за онлайн-курсы	Если бы Вы могли вновь принимать решение о поступлении в вуз, то что бы Вы выбрали?	Желающие поступить не в этот вуз или не на эту же программу, чаще выбирали оффлайн-курсы
3	Большинство за смешанный формат: 39% за оффлайн-курсы, 45% за дисциплины со смешанным обучением, 6% за онлайн-курсы	Оцените, Ваше отношение к этим элементам учебного процесса от -2 до 2, где «-2» - безусловно негативное, «2» - безусловно позитивное: - обязательное включение хотя бы одной онлайн-дисциплины или дисциплины со смешанным обучением (онлайн-лекции и оффлайн-семинары / оффлайн-экзамен) в свой учебный план	Те, кто оценил свое отношение к этому элементу как негативное, чаще выбирали оффлайн-курсы
4	За оффлайн и за смешанный формат: 40% за оффлайн-курсы, 40% за дисциплины со смешанным обучением, 13% за онлайн-курсы	1) Если бы Вы могли вновь принимать решение о поступлении в вуз, то что бы Вы выбрали? 2) Оцените, Ваше отношение к этим элементам учебного процесса: - включение майнора в свой учебный план, - обязательное включение хотя бы одной онлайн-дисциплины или дисциплины со смешанным обучением (онлайн-лекции и оффлайн-семинары / оффлайн-экзамен) в свой учебный план, - проверка текстов курсовых и выпускных квалификационных работ на процент заимствований через систему LMS (проверка на плагиат) 3) Оцените, насколько Вы согласны с высказыванием: - в университете я ощущаю себя частью коллектива - преподавателям есть дело до моих достижений в учебе	Среди выбравших онлайн-формат больше желание сменить вуз или программу или вообще не поступать, негативное отношение к необходимости включать майноры в свои учебные планы, более положительное отношение к включению онлайн и смешанных дисциплин, хуже отношение к необходимости проверки на плагиат, хуже показатели адаптации к университету (к коллективу и к преподавателям)

Продолжение Таблицы А.4

№ кла-стера	Предпочтительный формат курса	Вопросы, повлиявшие на предпочтения относительно перевода лекций в онлайн формат	Ответы, повлиявшие на предпочтения относительно перевода лекций в онлайн формат
5	Большинство за оф-флайн: 46% за оффлайн-курсы, 39% за дисциплины со смешанным обучением, 9% за онлайн-курсы	<p>1) Если бы Вы могли вновь принимать решение о поступлении в вуз, то что бы Вы выбрали?</p> <p>2) Оцените, Ваше отношение к этим элементам учебного процесса:</p> <ul style="list-style-type: none"> <li>- обязательное включение хотя бы одной онлайн-дисциплины или дисциплины со смешанным обучением (онлайн-лекции и оффлайн-семинары / оффлайн-экзамен) в свой учебный план,</li> <li>- обязательное наличие в учебных планах студентов дисциплин или модулей дисциплин, ориентированных на формирование компетенций в области информационных технологий и работы с данными (Data Culture)</li> </ul> <p>3) Сталкивались ли Вы с трудностями во время учебного процесса, связанного с онлайн-дисциплинами?</p> <ul style="list-style-type: none"> <li>- мне пришлось самостоятельно оплатить прохождение онлайн-дисциплины, которую я включал(а) в свой учебный план</li> </ul> <p>4) Оцените, насколько Вы согласны с высказыванием:</p> <ul style="list-style-type: none"> <li>- в университете я ощущаю себя частью коллектива</li> <li>- преподавателям есть дело до моих достижений в учебе</li> </ul>	<p>Среди выбравших онлайн-формат больше желание сменить вуз или программу или вообще не поступать, более положительное отношение к включению онлайн и смешанных дисциплин, лучше отношение к обязательному включению дисциплин по использованию информационных технологий (особенно среди тех, кто за смешанный формат), хуже показатели адаптации к университету (к коллективу и к преподавателям). Среди слушателей курсов смешанного формата больше тех, кому пришлось столкнуться с необходимостью самостоятельной оплаты онлайн-курса</p>
6	За оффлайн и за смешанный формат: 41% за оффлайн-курсы, 39% за дисциплины со смешанным обучением, 8% за онлайн-курсы	<p>1) Оцените, Ваше отношение к этим элементам учебного процесса:</p> <ul style="list-style-type: none"> <li>- обязательное включение хотя бы одной онлайн-дисциплины или дисциплины со смешанным обучением (онлайн-лекции и оффлайн-семинары / оффлайн-экзамен) в свой учебный план.</li> </ul> <p>2) Оцените, насколько Вы согласны с высказыванием:</p> <ul style="list-style-type: none"> <li>- я чувствую себя комфортно среди одногруппников</li> </ul>	<p>Среди выбравших онлайн-формат более положительное отношение к включению онлайн и смешанных дисциплин, а среди выбравших смешанный формат хуже самочувствие среди одногруппников</p>

Продолжение Таблицы А.4

№ кла-стера	Предпочтительный формат курса	Вопросы, повлиявшие на предпочтения относительно перевода лекций в онлайн формат	Ответы, повлиявшие на предпочтения относительно перевода лекций в онлайн формат
7	Большинство за смешанный формат: 37% за оффлайн-курсы, 47% за дисциплины со смешанным обучением, 12% за онлайн-курсы	<p>1) Если бы Вы могли вновь принимать решение о поступлении в вуз, то что бы Вы выбрали?</p> <p>2) Оцените, Ваше отношение к этим элементам учебного процесса:</p> <ul style="list-style-type: none"> <li>- обязательное включение хотя бы одной онлайн-дисциплины или дисциплины со смешанным обучением (онлайн-лекции и оффлайн-семинары / оффлайн-экзамен) в свой учебный план,</li> </ul> <p>3) Оцените, насколько Вы согласны с высказыванием:</p> <ul style="list-style-type: none"> <li>- в университете я ощущаю себя частью коллектива</li> <li>- преподавателям есть дело до моих достижений в учебе</li> <li>- мои ожидания от НИУ ВШЭ оправдались</li> <li>- я чувствую себя безопасно в зданиях НИУ ВШЭ</li> </ul>	Среди выбравших онлайн-формат больше желание сменить вуз или программу или вообще не поступать, более положительное отношение к включению онлайн и смешанных дисциплин, хуже показатели адаптации к университету (нет ощущения себя частью коллектива, внимания от преподавателей, не оправдались ожидания от вуза и меньше ощущения безопасности в зданиях НИУ ВШЭ).

## ПРИЛОЖЕНИЕ Б

### Таблицы к разделу 5

Таблица Б.1 – Общая численность населения и изменение численности населения в Арктике за 1990-2018 гг. Источники: Norgregio, Census, Statistics Canada, Statistics Norway, Statistics Sweden, Statistics Finland, Rosstat, 2019

Total population and population change in the Arctic 1990 - 2018	Total population		Population change 1990-2018		
	1990	2018	Total	Natural increase	Migration
World	5 327 231 061	7 631 091 040	43,25	43.25	0.0
The USA	252 529 950	322 179 605	27,58	16.6	10.9
Alaska	553 171	737 080	33,25	37.0	-3.8
Canada	27 692 680	36 624 199	32,25	31.2	1,05
Yukon	27 797	35 874	29,06	15.0	14.0
Northwest Territories	40 845	44 597	9,19	9.0	0.2
Nunavut	27 498	37 996	38,18	41.4	-6.1
Greenland	55 558	55 877	0,57	24.8	-23.1
Iceland	253 785	348 450	38	27.2	10.0
Faroe Islands	47 773	50 498	5,7	17.4	-11.6
Norway	4 233 116	5 295 619	25,1	10.7	14.4
Nordland	239 532	243 335	1,59	4.1	-2.5
Troms	146 591	166 500	13,58	11.9	1.6
Finnmark	74 148	76 167	2,72	12.2	-9.5
Svalbard	3 544	2 637	-25,59	N/A	N/A
Finland	4 974 383	5 513 130	10,83	52	5.5
Lappi	199 973	179 223	-10,38	2.1	-12.5
North Ostabortten	348 292	411 856	18,25	18.6	-0.3
Kainuu	92 458	73 960	-20,01	-2,9	-172
Sweden	8 527 036	10 120 242	18,68	4.8	13.9
Vasterbotten	250 134	268 465	7,33	2.7	4.7
Norrbotten	262 839	251 295	-4,39	-1.5	-2.9
Russia	148 041 641	146 880 431	-0,78	-8,7	6.5
Republic of Karelia	791 720	627 000	-20,81	-12.5	-8.3
Komi Republic	1 248 891	850 000	-31,94	-1.9	-30.1
Arkhangelsk Oblast	1 575 502	1 166 000	-25,99	-8.5	-17.5

## Продолжение Таблицы Б.1

Total population and population change in the Arctic 1990 - 2018	Total population		Population change 1990-2018		
	1990	2018	Total	Natural increase	Migration
Murmansk Oblast	1 191 458	757 000	-36,46	-2.3	-34.1
Khanty-Mansi Aut. Okrug	1 267 030	1 646 000	29,91	23.1	6.8
Yamalo-Nenets Aut. Okrug	489 161	536 000	9,58	24.8	-15.2
Taymyr Aut. Okrug	51 867	34 432	-33,61	7.8	-41.4
Sakha Republic (Yakutia)	1 111 480	963 000	-13,36	15.5	-28.8
Chukotka Aut. Okrug	162 135	50 000	-69,16	4.3	-73.5
Kamchatka Oblast	476 911	315 000	-33,95	0.5	-34.4
Magadan Oblast	390 276	146 000	-62,59	-0.3	-62.3

Таблица Б.2 – Водообеспеченность стран Африки на душу населения

Не испытывающие водного дефицита (>1700 m <sup>3</sup> /year/capita)	
Страна	Водообеспеченность на душу населения
Ангола	4 983
Бенин	2 361
Ботсвана	5 340
Камерун	11 769
Центральноафриканская республика	30 264
Чад	3 067
Республика Конго	158 145
Кот-д'Ивуар	3 463
Демократическая республика Конго	15 773
Экваториальная Гвинея	20 505
Эсватини	3 299
Габон	81 975
Гамбия	3 808
Гана	1 949
Гвинея	17 771
Гвинея-Бисау	16 873
Либерия	49 028
Мадагаскар	13 179
Мали	6 472
Мавритания	2 579
Маврикий	2 175
Мозамбик	7 317
Намибия	15 750
Сан-Томе и Принсипи	10 671
Сенегал	2 459
Сьерра-Леоне	21 172
Южный Судан	3 936

Продолжение Таблицы Б.2

Не испытывающие водного дефицита (>1700 m <sup>3</sup> /year/capita)	
Страна	Водообеспеченность на душу населения
Того	1 885
Замбия	6 131
Испытывающие водный стресс (1000–1700 m <sup>3</sup> /year/capita)	
Страна	Водообеспеченность на душу населения
Бурунди	1 154
Коморские острова	1 474
Эритрея	1 443
Эфиопия	1 162
Лесото	1 353
Нигер	1 585
Нигерия	1 499
Руанда	1 089
Уганда	1 402
Танзания	1 680
Зимбабве	1 210
Испытывающие хронический дефицит водных ресурсов (500–1000 m <sup>3</sup> /year/capita)	
Страна	Водообеспеченность на душу населения
Буркина-Фасо	703.4
Кабо-Верде	549
Египет	589.4
Кения	617.7
Малави	927.9
Марокко	811.4
Южно-Африканская Республика	905.4
Судан	932.6
Испытывающие абсолютный дефицит воды (<500 m <sup>3</sup> /year/capita)	
Страна	Водообеспеченность на душу населения
Алжир	282.4
Джибути	313.5
Ливия	109.8
Тунис	400.2
Отсутствующие данные	
Страна	Водообеспеченность на душу населения
Сейшельские острова	-
Реюньон	-
Майотта	-
Сомали	-

Примечание - Источник: AQUASTAT database n.d., (retrieved 29 November 2019, <<http://www.fao.org/nr/water/aquastat/data/query/index.html?lang=en>>)