

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «МОСКОВСКИЙ
ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ (НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)»
КАФЕДРА МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ СЛОЖНЫХ СИСТЕМ
И ОПТИМИЗАЦИИ

На правах рукописи

Козицин Иван Владимирович

**ВАЛИДАЦИЯ МОДЕЛИ КРАСНОЩЕКОВА НА
ОСНОВАНИИ ДАННЫХ ИЗ ОНЛАЙНОВОЙ
СОЦИАЛЬНОЙ СЕТИ ВКОНТАКТЕ**

Специальность 05.13.18 —

«Математическое моделирование, численные методы и комплексы программ»

Диссертация на соискание учёной степени
кандидата физико-математических наук

Научный руководитель:

д. ф.-м. н., проф.

Белолипецкий Александр Алексеевич, главный научный сотрудник
Федерального исследовательского центра «Информатика и управление»
Российской Академии Наук

Долгопрудный — 2019

Оглавление

	Стр.
Введение	4
Глава 1. Модели формирования мнений	12
1.1 Классификация и общая характеристика моделей формирования мнений	12
1.2 Некоторые выпуклые модели формирования мнений	22
1.2.1 Модель ДеГроота	22
1.2.2 Модель Фридкина-Джонсона	27
1.2.3 Протокол спорщиков	31
1.2.4 Задача восстановления сети влияния для модели Фридкина-Джонсона	32
1.2.5 Модель Краснощекова	34
1.3 Проблема расщепления общества и феномен поляризации	36
Глава 2. Анализ модели Краснощекова	38
2.1 Случай разреженной матрицы относительных влияний	38
2.2 Связь с моделями ДеГроота и Фридкина-Джонсона	45
2.3 Сходимость мнений и достижимость консенсуса	47
2.4 Сходимость мнений и достижимость консенсуса: случай существования изолированных конформистов, влияющих на коллектив	49
2.5 Разрешение временного парадокса в модели Краснощекова	57
2.6 Модель Краснощекова: расщепление общества и поляризация мнений	58
Глава 3. Анализ динамики мнений пользователей социальной сети ВКонтакте	60
3.1 Краткий обзор методологии	60
3.2 Предсказательная модель для определения политических взглядов пользователей ВКонтакте	61
3.2.1 Датасеты	62

	Стр.
3.2.2 Признаки	65
3.2.3 Выгрузка данных из Вконтакте	66
3.2.4 Разведочный анализ данных	67
3.2.5 Кодировка	69
3.2.6 Процедура отбора классификаторов	71
3.2.7 Отсеивание аномалий	74
3.2.8 Балансировка классов	76
3.2.9 Понижение размерности	76
3.2.10 Классификатор	77
3.2.11 Влияние выбора комбинации признаков на качество классификации	79
3.2.12 Анализ Д2	80
3.2.13 Построение одномерного пространства мнений	82
3.3 Данные о динамике мнений пользователей и их разведочный анализ	83
3.4 Аппроксимирующие зависимости	87
3.5 Дизайн экспериментов	90
3.6 Результаты аппроксимации	92
3.7 Анализ результатов	93
Заключение	102
Список сокращений и условных обозначений	103
Словарь терминов	104
Список литературы	106
Список рисунков	115
Список таблиц	117
Приложение А. Численности групп пользователей	118

Введение

Степень ее разработанности.

Мнение человека, его взгляды, привычки, убеждения и даже поступки имеют тенденцию меняться при взаимодействии с окружающим социумом, что вызвано действием социального влияния — силы, направленной на индивида со стороны:

- окружающих его людей;
- различных информационных источников (книги, газеты, радио, телевизор, Интернет);
- культурных норм;
- и др.

Опыты Шерифа [1] и Аша [2], проведенные в первой половине и в середине двадцатого века, соответственно, являются одними из первых экспериментальных подтверждений того факта, что человек пытается снизить уровень разногласий с окружающими его людьми, регулируя свое поведение и даже взгляды соответствующим образом. Этот феномен принято называть конформизмом [3]. Конформизм является одной из сторон социального влияния, однако возможны и другие его вариации [4].

Указанные выше эксперименты послужили фундаментом для первых формальных моделей, описывающих процесс формирования мнения индивидом в обществе (далее для обозначения таких моделей будут использоваться синонимичные термины "модель социального влияния", "модель формирования мнений" и "модель динамики взглядов"). Вся вторая половина двадцатого века ознаменовалась попытками формализовать, как социальное влияние формирует мнение человека. К настоящему моменту было предложено огромное количество моделей динамики взглядов, и, более того, их число продолжает неуклонно расти [5]. Одной из причин этому является тот факт, что социальное влияние имеет множество различных проявлений, при этом каждое из них можно формализовать по-разному, утилизировав разнообразные математические соотношения и в результате получая различные модели. Как правило, авторы данных исследований пропускают шаг обоснования выбора того или иного способа формализации. Таким образом, имея под собой только эмпирический фундамент из исследований в области социальной психологии (сюда, в частности, относятся

упомянутые выше эксперименты Шерифа и Аша), большинство исследований моделей динамики взглядов носят преимущественно теоретический характер [4–7]. Одной из причин этому является высокая сложность задачи сбора реальных данных, которые необходимы для этапа валидации модели. Особенно это касается микро-моделей, или агентных моделей, – моделей, описывающих динамику мнения каждого индивида по отдельности [5; 8]. Для таких моделей, даже в случае проведения подконтрольных экспериментов в лабораторных условиях, огромной проблемой является набрать достаточное число испытуемых [7; 9].

Развитие Интернета в целом и онлайн-социальных сетей (далее – ОСС), таких как Facebook, Twitter или ВКонтакте, в частности, подстегнуло интерес к моделированию социального влияния [4; 5; 10–12]. С распространением этих технологий изменилась скорость передачи информации, а также уровень ее доступности – человек в Интернете оказался окружен потоками данных, полностью обработать которые он попросту не в состоянии – необходимо делать выбор [11; 13; 14]. Важным обстоятельством является тот факт, что зачастую эта информация носит заведомо ложный характер – выпуклым примером являются так называемые fake news [15]. Теперь, при построении моделей динамики мнений необходимо учитывать эти, а также многие другие факторы. Здесь необходимо отметить, что ОСС являются проводниками не только информации, но также и социального влияния [16].

Пользователи, которые используют ОСС для общения и получения информации, генерируют при этом огромное количество данных, находящихся в открытом доступе. Согласно статистике (источник: <https://www.osp.ru/os/2013/08/13037856/>), в 2013 году Twitter и Facebook генерировали ежедневно 7 и 10 Тбайт данных соответственно. Эта информация потенциально может быть использована для проверки адекватности тех или иных моделей формирования мнений, поскольку позволяет восстанавливать некоторые скрытые характеристики самих пользователей, фигурирующие в уравнениях, описывающих динамику мнений в этих моделях [4–6; 17]. В частности, было установлено, что политические взгляды политически ангажированного пользователя могут быть определены с достаточно высокой точностью на основании информации из его аккаунта в ОСС [18–23]. Кроме того, полезной для анализа является информация о том, как пользователи связаны друг с другом функционалом соответствующей платформы. В частности, в ОСС ВКонтакте пользователи могут подписываться друг на друга, а если два пользователя подписаны друг на друга,

то они становятся “друзьями”. При этом, данные связи видны окружающим (в случае отсутствия настроек приватности). Согласно некоторым исследованиям, сила этих связей также может быть рассчитана (проранжирована) на основании данных из ОСС [16]. Таким образом, такие ключевые параметры моделей динамики мнений, как

- мнения пользователей;
- структура их социальных взаимодействий;
- сила этих взаимодействий

могут быть оценены на основании открытых данных из ОСС.

Однако, при работе с ОСС возникают определенные трудности: например, не все люди используют ОСС, в то время как даже те, кто это делает, могут скрывать свои аккаунты настройками приватности. Важным препятствием является смещенность пользователей по возрастному признаку в сторону более молодого населения [24]. Кроме того, обитателями ОСС являются не только обычные пользователи, но и искусственные аккаунты — боты — преследующие различные цели [11]. Эти же цели могут преследовать обычные пользователи — так называемые виртуалы, — которых наняли за деньги для выполнения тех или иных функций [11]. В результате, мы получаем частичную, искаженную картину, которая может сильно отличаться от реальной ситуации. Тем не менее, использование столь богатого источника информации, как ОСС, при условии правильной интерпретации результатов, учитывающей описанные выше факторы, позволит анализировать траектории мнений не малого количества испытуемых, как было до этого в лабораторных экспериментах, а сразу нескольких тысяч или даже миллионов пользователей, что значительно повысит репрезентативность данных и даст возможность строить статистически более надежные выводы [25]. Таким образом, проблема правильного использования больших данных из ОСС с целью проверки адекватности моделей социального влияния является наиболее актуальной в данном научном направлении [4—7]. При аккуратном использовании этой информации можно подтвердить/опровергнуть/уточнить те или иные модели социального влияния и механизмы формирования мнений.

Кроме того, необходимо учитывать тот факт, что в настоящее время ОСС являются местом информационных войн, происходящих на самых различных уровнях, в том числе на межгосударственном. Благодаря более глубокому пониманию того, как формируется мнение человека и, в частности, пользователя, можно:

- обеспечить более плодотворное планирование политических и маркетинговых кампаний;
- увеличить степень политической стабильности в стране;
- повысить эффективность управления в коммерческих организациях и других структурах;
- улучшить существующие методы по борьбе с экстремизмом.

Все это в совокупности определяет актуальность темы диссертационной работы.

Целью данной работы является проверка адекватности модели Краснощекова на основании данных из ОСС ВКонтакте.

Для достижения поставленной цели необходимо было решить следующие **задачи**:

1. Разработать обобщение модели Краснощекова на случай разреженной матрицы относительных влияний;
2. Создать математические методы и модели оценивания политических взглядов российских пользователей ОСС ВКонтакте, разработать алгоритм и соответствующее программное обеспечение для оценивания политических взглядов российских пользователей ОСС ВКонтакте;
3. Провести конструирование временного ряда, описывающего динамику мнений выборки российских пользователей ОСС ВКонтакте; зафиксировать структуру графа дружбы, соединяющего этих пользователей;
4. Исследовать полученные данные с точки зрения модели Краснощекова и, возможно, некоторых других моделей динамики мнений;
5. Предложить поправки для модели Краснощекова и, возможно, некоторых других моделей динамики мнений на основании полученных в предыдущем пункте результатов, а также определить направления для дальнейших исследований в данной области.

Научная новизна:

1. Впервые разработано обобщение модели Краснощекова на случай разреженной матрицы относительных влияний, исследованы свойства полученной модели;
2. Впервые найдена связь между моделью Краснощекова с одной стороны и моделями ДеГроота и Фридкина-Джонсона с другой;
3. Впервые исследованы асимптотические свойства модели Краснощекова;

4. Впервые сконструирован временной ряд, описывающий динамику мнений выборки российских пользователей ОСС ВКонтакте, состоящей более чем из миллиона аккаунтов, а также граф, фиксирующий дружеские связи этих пользователей;
5. Впервые проведен анализ адекватности упрощенных версий моделей Краснощекова и ДеГроота и получены экспериментальные подтверждения выпуклому механизму формирования мнений на выборке, состоящей более чем из полутора миллионов аккаунтов пользователей ОСС ВКонтакте;
6. Впервые предложены поправки для моделей Краснощекова и ДеГроота, учитывающие результаты натурного эксперимента подобного масштаба.

Практическая значимость работы заключается в том, что знание закономерностей, лежащих в основе механизмов формирования мнений могут помочь в:

1. Прогнозировании результатов выборов;
2. Планировании политических и маркетинговых кампаний;
3. Улучшении существующих методов по борьбе с экстремизмом;
4. Противодействию информационным атакам со стороны других государств.

Кроме того, разработанные модель, алгоритм и соответствующее программное обеспечение для оценивания политических взглядов российских пользователей ВКонтакте могут служить полезным инструментом для оценивания текущего состояния общественного мнения.

Методология и методы исследования.

Методологическую базу исследования составили такие модели динамики мнений, как модель ДеГроота, модель Фридкина-Джонсона и модель Краснощекова, а также результаты из теории неотрицательных матриц и теории графов.

Для проведения натурного эксперимента по проверке адекватности моделей Краснощекова и ДеГроота были задействованы данные из ОСС ВКонтакте, которые выгружались с помощью методов VK API (VKontakte Application Programming Interface). Полученная информация обрабатывалась методами интеллектуального анализа данных (в частности, для разработки предсказательной модели оценивания политических взглядов пользователей ВКонтакте были использованы подходы из теории машинного обучения). Для анализа графовых

структур применялись подходы из теории анализа сложных сетей и, в частности, социальных сетей.

Основные положения, выносимые на защиту:

1. Построено обобщение модели Краснощекова на случай разреженной матрицы относительных влияний, исследована связь между моделью Краснощекова с одной стороны и моделями ДеГроота и Фридкина-Джонсона с другой, изучены асимптотические свойства модели Краснощекова, с помощью протокола спорщиков разрешен временной парадокс модели Краснощекова;
2. Разработаны модель, алгоритм и соответствующее программное обеспечение для оценивания политических взглядов российских пользователей ВКонтакте;
3. Сконструирован временной ряд, описывающий динамику мнений выборки российских пользователей ВКонтакте, а также граф, фиксирующий дружеские связи этих пользователей;
4. Полученные данные исследованы с точки зрения упрощенных версий моделей Краснощекова и ДеГроота на предмет адекватности последних, обнаружены паттерны социального влияния, получено экспериментальное подтверждение явления конформизма в целом и данных моделей в частности с некоторыми дополнительными оговорками;
5. Предложены поправки для моделей Краснощекова и ДеГроота, учитывающие результаты натурного эксперимента, определены направления для дальнейших исследований в данной области.

Достоверность полученных результатов обеспечивается корректностью математических доказательств и выводов, тщательно разработанным дизайном натурного эксперимента для проверки адекватности моделей, аккуратностью и точностью его проведения, а также доскональностью анализа результатов эксперимента.

Апробация работы. Основные результаты работы докладывались на:

- VIII Московской международной конференции по Исследованию Операций (ORM-2016, Москва)
- Научном семинаре кафедры высшей математики МФТИ (2016, Долгопрудный)
- Семинаре по теории управления организационными системами (2016, ИПУ РАН, Москва)

- Семинаре лаборатории Наук о больших данных и проблемах общества, ТГУ с участием Натальи Касперской и Игоря Ашманова (2017, Томск)
- Ежегодных всероссийских конференциях Город-IT (2017, 2018, Томск)
- Семинарах по социофизике имени Д.С.Чернавского (2018, МГУ и 2019, ИПУ РАН, Москва)
- Второй всероссийской междисциплинарной конференции «СОЦИОФИЗИКА И СОЦИОИНЖЕНЕРИЯ» (2018, Москва)
- IX Московской международной конференции по Исследованию Операций (ORM-2018, Москва)
- Одиннадцатой международной конференции «Управление развитием крупномасштабных систем» (MLSD 2018, Москва)
- XIII Всероссийском совещании по проблемам управления (2019, Москва)
- Двенадцатой международной конференции «Управление развитием крупномасштабных систем» (MLSD 2019, Москва)

Личный вклад. Все результаты диссертации были получены автором самостоятельно, включая результаты, опубликованные в работах с соавторами.

Публикации. Основные результаты по теме диссертации изложены в 15 публикациях, 4 из которых индексируются Web of Science и Scopus.

Объем и структура работы. Диссертация состоит из введения, трёх глав, заключения и одного приложения.

В первой главе приведены классификация и общая характеристика моделей формирования мнений. Кроме того, первая глава посвящена описанию некоторых выпуклых синхронно-итеративных моделей динамики мнений – ДеГроота, Фридкина-Джонсона и Краснощекова, – а также их различным модификациям и некоторым сопутствующим вопросам.

Во второй главе строится обобщение модели Краснощекова на случай разложимой матрицы относительных влияний, исследуется связь между моделью Краснощекова с одной стороны и моделями ДеГроота и Фридкина-Джонсона с другой, анализируется вопрос сходимости мнений в модели Краснощекова. Там же разрешается временной парадокс модели Краснощекова.

Третья глава посвящена анализу динамики мнений пользователей ОСС ВКонтакте. В ней приведены модель, алгоритм и комплекс программ для оценивания политических предпочтений российских пользователей ОСС ВКонтакте. Там же представлен временной ряд, описывающий динамику мнений выборки

пользователей ВКонтакте, а также граф, фиксирующий структуру соединяющих их дружеских связей. В рамках третьей главы полученные данные анализируются через призму моделей Краснощекова и ДеГроота, разрабатываются поправки к этим моделям, учитывающие результаты анализа данных натурального эксперимента.

В заключении сформулированы основные результаты.

В приложении приведена вспомогательная статистика.

Полный объём диссертации составляет 118 страниц, включая 23 рисунка и 7 таблиц. Список литературы содержит 98 наименований.

Глава 1. Модели формирования мнений

1.1 Классификация и общая характеристика моделей формирования мнений

Модели формирования мнений описывают, как меняются мнения и/или поведение людей под действием социального влияния, вызванного их взаимодействием с окружающим социумом [4—7; 26; 27]. Понятия “мнение” и “поведение” не являются синонимами, однако тесно связаны между собой, поскольку первое является предтечей второго [7]. Кроме того, важным обстоятельством является тот факт, что в рамках социальных контактов, люди напрямую не видят мнения друг друга, а лишь получают информацию, сформированную в результате действий их собеседников [28]. Стоит отметить, что многие модели формирования мнений в целях простоты не различают эти две сущности. Однако, некоторые, более современные, модели учитывают, что поступки людей могут расходиться с их реальными позициями [28].

Модели формирования мнений описывают, как под действием социального влияния изменяются [4]:

- отношение индивида к другому человеку, событию, действию и т.д. [5];
- ответ на некоторый вопрос, например: за какую партию индивид пойдет голосовать на выборах [29; 30];
- степень уверенности индивида в конкретном поступке [31];
- вероятность, что индивид совершит конкретный поступок [32];
- и др.

Перечисленные выше пункты, будучи тесно переплетенными между собой, показывают, что различные модели динамики мнений могут давать различную интерпретацию понятию “мнение”, однако, при формализации этой сущности (см. далее в этом параграфе) эти различия существенным образом сглаживаются.

Перед тем как перейти к понятию социального влияния, логичным будет упомянуть чрезвычайно простое определение понятия “мнение”, данное известным ученым Робертом Аксельродом [33]: мнение — это то, на что влияет социальное влияние. Далее, в качестве синонимов понятия “мнение” будут

использоваться термины “взгляды” и “убеждения”. При этом, если не оговорено обратное, будет предполагаться, что мнение человека и его поведение совпадают.

Как уже было сказано во введении, социальное влияние – это многогранная сущность, каждое из проявлений которой представляет огромный интерес. Существуют три возможных исхода действия социального влияния на индивида (объекта влияния):

1. отсутствие каких-либо изменений мнения/поведения индивида;
2. ассимилятивный исход (положительное влияние) – мнение/поведение индивида стало ближе ко мнению/поведению источника (субъекта) социального влияния;
3. репульсивный исход (отрицательное/негативное влияние) – мнение/поведение объекта влияния отделилось от мнения субъекта влияния.

Среди перечисленных выше, наибольшее внимание научного сообщества получил ассимилятивный исход, являющийся проявлением конформизма [32]. Более того, если сравнивать второй и третий пункты (соответствующие тому, что мнение объекта влияния изменилось), то именно положительное социальное влияние в наибольшей степени распространено и эмпирически подтверждено [1; 2; 4; 9; 16; 34]. Причиной ему могут служить:

- действие убеждающих аргументов (которые могут быть ложными, или, в некоторых случаях, давать неполную картину мира) [35];
- недостаток информированности объекта влияния, заставляющий его следовать мнению/поведению того, кто, как кажется фокальному индивиду, владеет информацией [36];
- желание влиться в коллектив – социализация [37];
- подражание тому, кого объект влияния уважает и на кого он хочет быть похожим [36];
- нежелание идти на конфронтацию [38];
- и др.

Что касается негативного влияния, основным объяснением ему служит предположение, что если два индивида сильно отличаются друг от друга (их мнения/поведение, или же иные характеристики, например, цвет кожи, религия имеют существенные отличия), то результатом их общения будет рост различий между их мнениями/поведением [39]. Этот тезис является производной от другого, согласно которому положительное влияние может иметь место только в

том случае, если субъект влияния и объект влияния не сильно отличаются друг от друга - эффект ограниченного доверия [40].

Стоит отметить, что существующий эмпирический базис, обосновывающий существование негативного влияния подвержен критике [4]. Тем не менее, модели социального влияния, учитывающие наличие репульсивных взаимодействий (например, модель Альтафини [41]) способны объяснить явление поляризации, являющееся сигнатурой нашего времени (см. параграф 1.3).

Существуют два основных подхода к моделированию динамики мнений – *микроподход* и *макроподход* [5; 42].

В рамках микроподхода рассматривается система, состоящая из агентов (упрощенных двойников моделируемых людей, а также других действующих акторов, например – источников пропаганды), каждый из которых имеет собственное мнение, а также некоторые другие характеристики, и связей между ними [5; 42]. При этом, микромодели описывают динамику мнения каждого агента – такие модели еще называют агентными [7].

В случае макромоделей речь идет об анализе некоторых агрегированных показателей, например, доли людей, разделяющих данное мнение (поддерживающих данную партию), не конкретизируя, что это за люди [43–45]. В связи с этим, валидация макро-моделей является относительно более простой задачей, поскольку необходимые для нее данные зачастую находятся в открытом доступе. Наиболее выпуклым примером здесь являются социальные опросы, проводимые различными организациями (к примеру, ФОМом и ВЦИОМом): информация, предоставляемая этими организациями, носит усредненный и обезличенный характер, поэтому в случае анализа динамики общественного мнения можно исследовать только изменения агрегированных показателей, в то время как изменения мнений отдельных людей остаются скрытыми.

Стоит отметить, что одна из первых известных работ, посвященных моделированию социального влияния, имеет в своей основе макроподход. Речь идет о модели Рашевского [46], соответствующая статья была опубликована в 1939 году. В рамках модели Рашевского рассматривается группа из n индивидов. Часть из них относятся к типу 1 (интерпретируемому как множество людей с сформированными радикальными взглядами). Среди индивидов первого типа существуют две диаметрально противоположные точки зрения. В зависимости от того, какой точки зрения придерживаются индивиды, они могут принадлежать множествам I_A или I_B с численностями $x_0 = |I_A|$ и $y_0 = |I_B|$ соответственно.

Остальные люди принадлежат к типу 2. Предполагается, что индивиды типа 1 влияют на людей типа 2, приглашая разделить их взгляды. Иными словами, идет борьба за мнения $n - x_0 - y_0$ сомневающихся людей. Пусть число людей типа 2 в момент времени t , которые под влиянием индивидов I_A (соответственно I_B) присоединились к взглядам последних, равно $x(t)$ (соответственно $y(t)$). В этом случае

$$x(t) + y(t) = n - x_0 - y_0.$$

Дифференциальные уравнения, описывающие динамику рассматриваемой системы, имеют следующий вид:

$$\begin{cases} \dot{x}(t) = a_0 x_0 + ax(t) - c_0 y_0 - cy(t); \\ \dot{y}(t) = c_0 y_0 + cy(t) - a_0 x_0 - ax(t). \end{cases} \quad (1.1)$$

В (1.1) параметры a , a_0 , c и c_0 – неотрицательные коэффициенты, фиксирующие силы влияния соответствующих групп. Сформулированная модель позволяет изучить, каким образом a , a_0 , c и c_0 , а также начальные условия x_0 и y_0 влияют на итоговое распределение мнений людей, задаваемое пределами $\lim_{n \rightarrow \infty} x(t)$ и $\lim_{n \rightarrow \infty} y(t)$.

Как уже было сказано, данная модель относится к классу макромоделей: описанные соотношения не позволяют установить, какую позицию выберет данный конкретный человек, но дают информацию о том, сколько индивидов будет иметь данное конкретное мнение в определенный момент времени.

В рамках настоящей диссертации значительное внимание уделено исследованию модели Краснощекова, которая относится к классу микромоделей, вследствие чего дальнейшее изложение будет ограничено именно этим классом моделей. Одним из первых таких исследований является модель Френча [47]. Дальнейшее развитие этой модели связано с работами Харари [48] и ДеГроота [49]. Близкими по смыслу являются модели Абельсона [50], Тейлора [51] и Фридкина-Джонсона [31]. Модель Краснощекова [32], изучению и валидации которой посвящена настоящая диссертация, также имеет много общего с указанными моделями. Относительно недавно увидела свет модель отраженных оценок (модель ДеГроота-Фридкина), описывающая не только динамику мнений, но и то, как меняется сеть влияния, связывающая агентов [52; 53]. Все эти работы связаны одной общей сигнатурой: в их основе лежит так называемый

выпуклый механизм формирования взглядов, согласно которому мнение агента в следующий момент времени (оцениваемое скаляром или вектором) есть выпуклая комбинация мнения этого же агента в предыдущий момент времени, а также мнений агентов, влияющих на него. Ввиду особой важности данного класса моделей, им будет уделен отдельный параграф (см. параграф 1.2).

Агентные модели формирования мнений делятся на байесовские и небайесовские. Первые, как следует из названия, основаны на теореме Байеса. В самом простом случае это выглядит следующим образом [26]. Пусть мнение агента i до взаимодействия с окружающим его социумом описывается распределением вероятностей $P\{\theta\}$. Для простоты можно считать, что речь идет о некоторой неизвестной агенту i величине $\theta \in \Theta$, значение которой он пытается оценить. При этом, распределение $P\{\theta\}$ выражает, какой вероятностью индивид оценивает равенство θ тому или иному значению $x \in \Theta$. В результате взаимодействия с социумом, агент i получает порцию информации $s \in S$. В следствие этого, представления индивида о величине θ меняются следующим образом:

$$P\{\theta = x|s\} = \frac{P\{\theta = x, s\}}{P\{s\}} = \frac{P\{s|\theta = x\} P\{\theta = x\}}{P\{s\}} \quad (1.2)$$

В (1.2) предполагается, что агент i знает величины $P\{s|\theta = x\}$, $P\{\theta = x\}$ и $P\{s\}$. Иными словами, у него в голове сформирована некая модель мира, согласно которой он вычисляет $P(\theta = x|s)$. Данную процедуру принято называть социальным обучением: индивид изменяет свое мнение на основании информации, поступающей извне [54].

В дальнейшем, в фокусе внимания будут небайесовские модели формирования мнений, к которым, в частности, относится модель Краснощекова. Как было отмечено во введении, число моделей динамики взглядов (даже если ограничиваться только агентными и небайесовскими) неуклонно растет, поэтому охватить все существующие на данный момент чрезвычайно тяжело. Далее будет дана их краткая классификация, а также приведены некоторые из них, появившиеся, преимущественно, в двадцатом веке. Для более подробного ознакомления см. обзорные работы [4–6; 12; 17; 26].

Имеется несколько ключевых признаков, на основании которых можно классифицировать агентные модели динамики взглядов. Первый – это способ формализации времени. Существуют дискретные модели [29; 31; 32; 49], рассматривающие динамику системы в конечное (в предельном случае – счетное)

число моментов времени, и непрерывные модели, в которых время — непрерывная величина [50; 51; 55]. При этом, у некоторых дискретных моделей есть непрерывные аналоги и, наоборот, примеры: модель ДеГроота (дискретное время) и Абельсона (непрерывное время), модель Тейлора (непрерывное время) и модель Фридкина-Джонсона (дискретное время). Многие модели с дискретным временем используют математический аппарат теории разностных уравнений, в то время как модели с непрерывным временем преимущественно оперируют дифференциальными уравнениями. К примеру, достаточно широкий класс механизмов формирования взглядов в случае дискретного времени может быть формализован с помощью следующего разностного уравнения:

$$p_i(t+1) = p_i(t) + \Delta p_i, \quad i \in [n]. \quad (1.3)$$

В (1.3) мнение агента i в момент времени t описывается скаляром $p_i(t) \in \mathbb{R}$, а изменение его мнения — величиной Δp_i , которая может зависеть от многих факторов, и, как правило, в их число входит мнения агентов, имеющих влияние на агента i в момент времени t . Далее множество этих агентов будет обозначаться как $V_i(t)$. В самом простом случае $\Delta p_i = f(\{p_j(t)\}_{j \in V_i})$, и тогда:

$$p_i(t+1) = p_i(t) + f(\{p_j(t)\}_{j \in V_i}), \quad i \in [n]. \quad (1.4)$$

Одним из наиболее простых для аналитических исследований и, в то же время, достаточно содержательным, является случай, когда функция f линейна:

$$f(\{p_j(t)\}_{j \in V_i}) = \sum_{j \in V_i} l_{ij} p_j(t), \quad i \in [n].$$

В таком случае (1.4) примет вид

$$f(\{p_j(t)\}_{j \in V_i}) = l_{ii} p_i(t) + \sum_{j \in V_i} l_{ij} p_j(t), \quad i \in [n]. \quad (1.5)$$

Величины l_{ij} в (1.5) (которые в самом общем случае могут принимать любые значения, в том числе и отрицательные) описывают, как агенты влияют друг на друга [5]. Эту информацию удобно представлять с помощью матрицы *влияний*.

Определение 1.1.1. Матрицей влияния для системы, описываемой уравнением (1.5) называется матрица $\mathbf{L} = [l_{ij}] \in \mathbb{R}^{n \times n}$.

Матрице влияний L соответствует *сеть влияний*.

Определение 1.1.2. *Сетью влияний, соответствующей матрице влияний $L \in \mathbb{R}^{n \times n}$, называется простой ориентированный взвешенный граф $G[L] = (V, E, L)$. Здесь V - множество вершин графа, соответствующее множеству агентов рассматриваемой системы ($|V| = n$), а E - множество ребер: $(j, i) \in E$ тогда и только тогда, когда $l_{ij} \neq 0$.*

Полагая $\mathbf{p}(t) = [p_1(t), \dots, p_n(t)]^T$, уравнение (1.5) можно переписать в матричной форме:

$$\mathbf{p}(t+1) = L\mathbf{p}(t). \quad (1.6)$$

Важным признаком для классификации моделей формирования мнений является то, каким образом представлены мнения агентов. В самых простых случаях мнение может выражать позицию индивида относительно одного конкретного вопроса. Тогда мнение представляется скаляром [29; 32]. Этот скаляр может быть дискретной величиной, принимающей конечное (счетное) число альтернатив – например, за какого кандидата голосовать на выборах [29; 30; 56], – или непрерывной, как в (1.6) [31; 32; 49]. Весьма известной интерпретацией дискретных величин, описывающих мнения, является сравнение со спинами частиц: исследователи социального влияния очень часто обращаются к модели Изинга из теории ферромагнетизма [57], берущей свое начало из статистической физики.

Крайне популярной и содержательной моделью, использующей дискретные мнения является **модель голосующего** [29; 30; 58]. В ее простейшей конфигурации рассматривается система из n агентов, мнения которых принимают одно из двух значений $s = \pm 1$. Задана сеть влияния, являющаяся, в отличие от сети влияния для системы (1.6), невзвешенным ориентированным графом $G = (V, E)$. Множество его вершин V соответствует агентам рассматриваемой системы ($|V| = N$), в то время как наличие ребра (j, i) от вершины j к вершине i означает, что агент j имеет влияние на агента i . Динамика мнений в модели голосующего подчининена следующему протоколу. В каждый момент времени t случайно выбранный агент i меняет свое мнение $s_i(t)$ на мнение случайно выбранного агента j среди тех, кто имеет влияние на агента i : $j \in V_i = \{k \in V \mid (k, i) \in E\}$.

Данная модель имеет множество различных обобщений, в том числе [7]:

– множество значений величины s_i состоит из более, чем двух величин;

- сеть влияния $G[W]$ является взвешенным графом: вероятности выбора субъекта влияния объектом влияния неравнозначны;
- агенты обладают разной восприимчивостью ко мнениям других людей: случайно выбранный объект влияния – агент i – с вероятностью $\mu_i \in [0, 1]$ может избежать воздействия субъекта влияния;
- и др.

Что касается континуальных множеств, задающих пространства мнений, их очень часто отождествляют с отрезком $[0, 1]$. Числа из этого отрезка выражают когнитивное отношение агента к данному вопросу/предмету/человеку или же действию [5]. Такой выбор пространства мнений позволяет насыщать модель вероятностными интерпретациями (одна из них – вероятность совершить некоторое действие индивидом) [32; 49], однако, при необходимости, можно выбрать любой другой промежуток за счет соответствующего преобразования прямой. Существуют, однако, и комбинированные подходы, учитывающие не только мнения, но и действия. К ним, в частности, относится CODA-модель (Continuous Opinions Discrete Actions), в рамках которой мнения описываются непрерывными величинами, а действия – дискретными [59].

Более сложная ситуация имеет место, если повестка дня включает в себя несколько вопросов для обсуждения — тогда мнение моделируется векторной величиной, размерность которой соответствует числу вопросов (или, к примеру, как в случае модели Аксельрода — культурных черт) [10; 31; 33]. При этом, вопросы могут быть логически независимы [10], а могут быть логически связаны [31; 60]. Здесь необходимо отметить работы [31; 60], где впервые была предложена модель (являющаяся обобщением модели Фридкина-Джонсона), учитывающая логическую связь между темами для обсуждений, а также, на примере опросов жителей США относительно войны в Ираке, было показано, каким образом логическая связь между топиками может влиять на итоговое распределение мнений.

Тенденция такова, что развитие моделей динамики мнений сопряжено с неизбежным усложнением описательных методов. Данное обстоятельство связано с желанием как можно более точно смоделировать такой сложный и многогранный процесс, как динамика человеческого мнения. Это касается и способов представления мнений: ранние модели оперировали с простейшими подходами (дискретные или непрерывные скаляры), в то время как современные используют все более сложные конструкции (векторы из логически связанных

величин). Однако не менее важно найти баланс между точностью и простотой модели.

Важнейшей составляющей любой модели динамики мнений является механизм формирования взглядов, то есть правило, на основании которого агенты меняют свои мнения. Согласно [4], практически все механизмы формирования взглядов могут быть отнесены к одному из трех классов:

1. Механизмы ассимиляционного влияния (конформизм): любое взаимодействие агентов друг с другом приводит к тому, что уровень разногласий между ними снижется (в крайних случаях – если агенты "упрямы" – не изменяется) [29; 31; 32; 49; 61].
2. Механизмы влияния схожести (ограниченное доверие): снижение уровня разногласий возможно, только если взаимодействуют агенты с похожими взглядами, иначе ничего не происходит [40; 62; 63].
3. Механизмы отталкивающего влияния: агенты, чьи мнения слишком сильно отличаются друг от друга, в результате взаимодействия могут увеличить уровень разногласий [41].

Из перечисленных выше, первые два пункта соответствуют ассимилятивному исходу действия социального влияния, в то время как третий относится к репульсивному.

Кроме того, имеются механизмы формирования взглядов, которые трудно однозначно отнести к одному из вышеперечисленных пунктов. Здесь следует упомянуть **USDF-модель** (United We Stand, Divided We Fall) [64]. В самом простом случае, она описывает динамику бинарных мнений агентов (мнение агента i описывается скалярной величиной $s_i \in \{-1, +1\}$), расположенных на одномерной решетке. В ее основе лежит предположение, что двум людям проще убедить человека, чем одному. Формализация данной идеи следующая. Рассмотрим четырех агентов, расположенных последовательно друг за другом. Без ограничения общности, их индексы $i-2$, $i-1$, i и $i+1$. Мнения агентов $i-1$ и i определяют мнения двух других агентов следующим образом: если $s_{i-1} = s_i$, то тогда $s_{i-2} = s_{i-1} = s_i = s_{i+1}$. В противном случае $s_{i-2} = s_i$ и $s_{i-1} = s_{i+1}$. В каждый момент времени случайно выбирается четверка расположенных друг за другом агентов, затем мнения двух крайних из них (самого левого и самого правого) преобразуются в соответствии с описанным протоколом.

При изучении динамики взглядов важным аспектом является структура социальных взаимодействий — то, каким образом агенты влияют друг на дру-

га. Эти связи принято формализовывать с помощью графа, который называют сетью влияния, и соответствующей ему матрицы – матрицы влияний (см. определения 1.1.1 и 1.1.2, как пример матрицы влияний и сети влияний для системы (1.5)). Вершинам этого графа соответствуют сами агенты. В свою очередь, ребра данного графа описывают то, как агенты влияют друг на друга. В самом простом случае сеть влияния – это простой ненаправленный (все связи взаимны) невзвешенный (все влияния равнозначны) граф. Более сложная конфигурация – простой взвешенный направленный граф, которая в состоянии более точно моделировать реальные социальные явления. Необходимо отметить, что сеть влияния не является графом дружбы. Тем не менее, общепринятой является гипотеза, что она должна иметь много общего с социальными графами, в том числе с графом дружбы. Стоит сказать, что даже задача идентификации такой относительно прозрачной конструкции, как граф дружбы, на основании эмпирических данных имеет множество трудностей. В свою очередь, сеть влияния – более сложный объект для изучения. В связи с этим, отдельное направление в теории моделей динамики мнений связано с задачей идентификации сети влияния [10]. Подробнее про это см. параграф 1.2.4

В силу того, что реальные сети влияния для исследователей недоступны, их зачастую имитируют с помощью тех или иных синтетических графов [4; 65]. Структура таких графов зависит от конкретной задачи и выбирается самим исследователем в зависимости от поставленной задачи. В самом простом случае это может быть регулярная решетка или полный граф, однако для достижения большего сходства с реальными социальными процессами используют более сложные модели синтетических графов, имитирующие те или иные свойства реальных социальных сетей [7], например модели Барбаси-Альберта [66] (для моделирования безмасштабных сетей) или Ваттса-Строгатца [67] (сетей с небольшим диаметром – small world networks).

Большинство существующих моделей динамики мнений описывают закрытую систему -- предполагается, что число агентов постоянно. Однако, в 2017 году вышла публикация [68], в которой рассмотрена модель, учитывающая возможность агентов приходить и покидать систему. Данное предположение определило новый вектор развития моделей социального влияния [68–70].

Существует еще множество свойств, относительно которых можно строить классификацию моделей формирования мнений. Тем не менее, перечисленные

выше составляют минимальный базис, позволяющий провести содержательный анализ.

1.2 Некоторые выпуклые модели формирования мнений

В данном параграфе будут приведены основные результаты касательно моделей ДеГроота и Фридкина-Джонсона.

1.2.1 Модель ДеГроота

В модели ДеГроота (в самой ее простой конфигурации) мнение агента i в момент времени t описывается скаляром $p_i(t) \in [a, b]$, где a, b – вещественные числа [47–49]. Без ограничения общности будем считать, что $a = 0, b = 1$. Динамика мнений агентов подчиняется уравнению (1.6), в котором матрица \mathbf{L} является строчно-стохастической. Чтобы отличать ее, будем использовать обозначение $\mathbf{W} = [w_{ij}] \in \mathbb{R}^{n \times n}$:

$$\mathbf{p}(t+1) = \mathbf{W}\mathbf{p}(t), \quad (1.7)$$

Величину $w_{ii} \in [0, 1]$ в (1.7) можно проинтерпретировать, как сопротивляемость агента i социальному влиянию: для данного i , чем больше элемент главной диагонали w_{ii} , тем меньше элементы w_{ij} вне главной диагонали, что следует из того, что матрица \mathbf{W} является строчно-стохастической. В силу этого, при больших значениях w_{ii} агент i обращает меньше внимания на мнения других агентов, сохраняя свои взгляды практически неизменными. Из строчной стохастичности матрицы \mathbf{W} также следует тот факт, что мнения агентов в следующий момент времени не выходят за пределы границ, задаваемых их же мнениями в предыдущий момент времени [5, с. 6]:

$$\begin{aligned} \max_i p_i(0) &\geq \max_i p_i(1) \geq \dots \geq \max_i p_i(t), \\ \min_i p_i(t) &\geq \dots \geq \min_i p_i(1) \geq \min_i p_i(0). \end{aligned} \quad (1.8)$$

Кроме того, имеет место линейная (и, следовательно, непрерывная) зависимость от начальных условий, поскольку мнения агентов в момент времени t могут быть получены из соотношения

$$\mathbf{p}(t) = \mathbf{W}^t \mathbf{p}(0). \quad (1.9)$$

Одним из центральных вопросов, связанных с анализом модели ДеГроота, является существование предела в (1.9) при $t \rightarrow \infty$ для любого начального распределения мнений, что эквивалентно существованию предела

$$\lim_{t \rightarrow \infty} \mathbf{W}^t, \quad (1.10)$$

то есть правильности матрицы \mathbf{W} [5]. Если предел в (1.10) существует, то матрица

$$\mathbf{W}^\infty = \lim_{t \rightarrow \infty} \mathbf{W}^t$$

связывает мнения агентов в начальный и "конечный" моменты времени:

$$\mathbf{p}(\infty) = \mathbf{W}^\infty \mathbf{p}(0).$$

Эту матрицу (которая также является строчно-стохастичной) в иностранной литературе принято называть *матрицей управления* (control matrix) [5]. Ее компоненты описывают, как итоговое распределение мнений зависит от начальных мнений агентов, иными словами выражают *социальную власть* (social power [47]) агентов друг над другом. Пусть

$$\mathbf{W}^\infty = \begin{bmatrix} w_{11}^\infty & \dots & w_{1n}^\infty \\ \dots & \dots & \dots \\ w_{n1}^\infty & \dots & w_{nn}^\infty \end{bmatrix}.$$

Тогда компонента w_{ij}^∞ описывает, какой вклад дает начальное мнение агента j в итоговое мнение агента i .

Помимо существования предела (1.10), важной задачей является понять, смогут ли агенты достичь *консенсуса*.

Определение 1.2.1. *Консенсусом называется такое распределение мнений агентов \mathbf{p} , при котором мнения агентов равны друг другу:*

$$\mathbf{p} = p \mathbb{1}_n.$$

Стоит отметить, что определение социальной власти, введенное в [47], подразумевает, что агенты при $t \rightarrow \infty$ достигли консенсуса (что эквивалентно регулярности матрицы \mathbf{W} – см. далее). В этом случае строки матрицы \mathbf{W}^∞ совпадают, и можно легко сравнивать вклады агентов в итоговое распределение мнений.

В силу линейности системы (1.7), вопросы существования предела (1.10) и достижимости консенсуса были подробнейшим образом изучены. Ответ на каждый из них может быть сформулирован как с точки зрения спектра матрицы \mathbf{W} , так и с точки зрения структуры графа $G[\mathbf{W}]$. Следующие две теоремы описывают критерий существования предела (1.10).

Теорема 1.2.1. [71] *Предел (1.10) существует тогда и только тогда, когда матрица \mathbf{W} является правильной.*

Теорема 1.2.2. [5, с. 6, теорема 12] *Предел (1.10) существует тогда и только тогда, когда все закрытые компоненты сильной связности графа $G[\mathbf{W}]$ аperiодичны.*

Простейшим примером системы, мнения агентов которой не сходятся, является сообщество из двух *конформистов* – агентов, чьи сопротивляемости социальному влиянию равны 0: $w_{ii} = 0$.

Пример 1.2.1. *Рассмотрим систему, задаваемую матрицей влияний*

$$\mathbf{W} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \quad (1.11)$$

Спектр матрицы (1.11) состоит из двух чисел: $\sigma(\mathbf{W}) = \{-1, 1\}$, каждое из которых лежит на единичной окружности в комплексной плоскости. При этом,

$$\mathbf{W}^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{W}^3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

и так далее, что означает, что эти два агента постоянно меняются мнениями друг с другом, препятствуя тем самым их сходимости. С другой стороны, граф $G[\mathbf{W}]$ содержит ровно одну закрытую компоненту сильной связности (сам являясь ей), которая является периодичной с периодом 2.

Критерии достижимости консенсуса имеют следующий вид.

Теорема 1.2.3. [71] Система (1.7) достигает консенсуса \mathbf{p} для любого начального распределения мнений тогда и только тогда, матрица \mathbf{W} является регулярной [5].

Теорема 1.2.4. [5, с. 6, теорема 12] Система (1.7) достигает консенсуса \mathbf{p} тогда и только тогда, когда все закрытые компоненты сильной связности графа $G[\mathbf{W}]$ апериодичны и их количество не превышает одну [5].

Примером системы, которая не в состоянии достичь консенсуса, может служить ситуация, когда в обществе представлены две влиятельные группы, изолированные от воздействий со стороны.

Пример 1.2.2. Рассмотрим систему, задаваемую матрицей влияний

$$\mathbf{W} = \begin{bmatrix} 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{bmatrix}. \quad (1.12)$$

Спектр матрицы (1.12) состоит из трех чисел: $\sigma(\mathbf{W}) = \{0, 0.4, 1\}$, при этом, единственный элемент спектра, лежащий на единичной окружности комплексного круга – это единица, следовательно, существует предел (1.10), который таков, что

$$\mathbf{W}^\infty = \begin{bmatrix} 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{bmatrix}.$$

В силу того, что строки \mathbf{W}^∞ не равны друг другу, предельное распределение мнений агентов не может быть консенсусом для любого начального условия $\mathbf{p}(0)$. С точки зрения структуры графа $G[\mathbf{W}]$, консенсусу препятствует наличие двух закрытых компонент сильной связности: первая состоит из одного агента $i = 3$, который полностью невосприимчив ко внешнему влиянию ($w_{ii} = 1$), в то время как вторая образована агентами $i = 4, 5$, которые влияют друг на друга, избегая при этом влияния со стороны других агентов.

Модель (1.7) естественным образом обобщается на случай, когда мнение агента относительно рассматриваемого (единственного) вопроса задается не одним числом, а распределением вероятностей [49]. Для случая, когда имеется конечное число исходов m , мнение агента i в момент времени t задается стохастической строкой $\mathbf{p}_i(t) = [p_{i1}(t), \dots, p_{im}(t)]$ (строкой, в которой все компоненты неотрицательны и в сумме дают единицу). Матрица $\mathbf{P}(t) \in \mathbb{R}^{n \times m}$, строками которой являются $\mathbf{p}_i(t)$, задает мнения всех агентов в момент времени t :

$$\mathbf{P}(t) = \begin{bmatrix} \mathbf{p}_1(t) \\ \dots \\ \mathbf{p}_n(t) \end{bmatrix}.$$

Произведение

$$\mathbf{W}\mathbf{P}(t) \in \mathbb{R}^{n \times m}$$

дает в результате матрицу той же размерности, что и $\mathbf{P}(t)$, все строки которой также являются стохастическими, что делает корректным использование уравнения

$$\mathbf{P}(t+1) = \mathbf{W}\mathbf{P}(t). \quad (1.13)$$

для описания динамики мнений агентов.

Еще одним обобщением (1.7) является рассмотрение ситуации, когда агенты обсуждают между собой не один топик, а m разных вопросов, которые логически **независимы** между собой [5]. В этом случае, мнение агента i в момент времени t также представляется строкой $\mathbf{p}_i(t) = [p_{i1}(t), \dots, p_{im}(t)]$, однако в этом случае она не является стохастической. Каждая компонента j этой строки отвечает за мнение агента i относительно топика j . Конкатенируя эти строки строки вместе одна над другой, получаем матрицу $\mathbf{P}(t) \in \mathbb{R}^{n \times m}$. Уравнение (1.13) подходит и для описания такой системы.

Отдельное направление исследований связано с определением скорости сходимости системы (1.7). С точки зрения теории управления, важными вопросами являются вычисление матрицы \mathbf{W}^∞ и поиск агентов с наибольшей социальной властью [11], а также выбор структуры матрицы \mathbf{W} , обеспечивающей наискорейшую сходимость к консенсусу [72].

Более сложной для изучения является система (1.7), в которой матрица \mathbf{W} не является постоянной, а зависит от некоторых величин, например от времени:

$\mathbf{W} = \mathbf{W}(t)$. Данный случай является более приближенным к реальным процессам, позволяя моделировать появления новых знакомств, а также изменение доверия людей друг к другу [73].

Большое внимание научного сообщества получила задача восстановления матрицы влияния \mathbf{W} на основании имеющихся наблюдений мнений агентов. Ее описанию уделен пункт 1.2.4.

1.2.2 Модель Фридкина-Джонсона

Система, описываемая соотношением (1.7), является Марковским процессом: мнения агентов в следующий момент времени $\mathbf{p}(t+1)$ зависят только от их же мнений в предыдущий момент $\mathbf{p}(t)$ через строчно-стохастическую матрицу. В связи с этим, агенты "лишены памяти". Модель Фридкина-Джонсона, напротив, учитывает тот факт, что мнения агентов могут на каждом шаге зависеть от каких-либо фундаментальных убеждений, *предубеждений* (prejudices), которые были заложены, к примеру, в ходе воспитания, взросления и т.д. Модель Фридкина-Джонсона [31; 74] описывает динамику мнений агентов с помощью соотношения

$$\mathbf{p}(t+1) = (\mathbf{I}_n - \mathbf{X})\mathbf{W}\mathbf{p}(t) + \mathbf{X}\mathbf{u}, \quad (1.14)$$

где $\mathbf{X} = \text{diag}(x_1, \dots, x_n)$. Ее компонента $x_i \in [0, 1]$ выражает приверженность агента i своим предубеждениям. Необходимо отметить, что в общем случае предполагается, что x_i и w_{ii} никак не связаны, однако иногда для простоты полагается, что $x_i = w_{ii}$.

Вектор $\mathbf{u} \in \mathbb{R}^n$ в (1.14) выражает предубеждения агентов и, вообще говоря, отличен от начальных мнений агентов $\mathbf{p}(0)$: $\mathbf{u} \neq \mathbf{p}(0)$. Тем не менее, в некоторых случаях предполагается, что это одна и та же величина.

Модель Фридкина-Джонсона является обобщением модели ДеГроота: последняя может быть получена из первой в предположении $x_i = 0$ для всех $i \in [n]$. Кроме того, тесная связь между этими двумя моделями обусловлена следующим результатом.

Теорема 1.2.5. Системе (1.14) поставим в соответствие величины $\hat{\mathbf{p}}(t) \in \mathbb{R}^{2N}$, $\hat{\mathbf{W}} \in \mathbb{R}^{2N \times 2N}$ и вектор $\hat{\mathbf{p}}(0) \in \mathbb{R}^{2N}$ такие, что

$$\hat{\mathbf{W}} = \begin{bmatrix} (\mathbf{I}_n - \mathbf{X})\mathbf{W} & \mathbf{X} \\ 0 & \mathbf{I}_n \end{bmatrix}, \quad \hat{\mathbf{p}}(0) = \begin{bmatrix} \mathbf{p}(0) \\ \mathbf{u} \end{bmatrix}. \quad (1.15)$$

и рассмотрим систему

$$\hat{\mathbf{p}}(t+1) = \hat{\mathbf{W}}\hat{\mathbf{p}}(t), \quad (1.16)$$

с начальным распределением мнений $\hat{\mathbf{p}}(0)$. Тогда для любого t выполнено равенство: $p_i(t) = \hat{p}_i(t)$, $i \in [n]$.

Теорема 1.2.5 утверждает, что динамику модели Фридкина-Джонсона можно симитировать с помощью модели ДеГроота за счет введения синтетических агентов, отвечающих за предубеждения изначальных агентов. При этом, синтетические агенты невосприимчивы ко внешнему влиянию, их мнения неизменны (так как предубеждения агентов в модели Фридкина-Джонсона предполагаются постоянными). Каждый из синтетических агентов имеет влияние x_i ровно на одного изначального агента, выражая тем самым предубеждения последнего.

Данная связь позволяет исследовать систему (1.14) с помощью результатов, полученных для модели ДеГроота, применяя их к (1.16). В частности, используя ее, легко получить критерий сходимости мнений для системы (1.14). Обозначим $\mathbf{B} = (\mathbf{I}_n - \mathbf{X})\mathbf{W}$.

Теорема 1.2.6. [5, с. 11, теорема 21] Мнения агентов в (1.14) сходятся тогда и только тогда, когда матрица \mathbf{B} является правильной или ее Фробениусово число меньше единицы.

Случай, когда матрица \mathbf{B} является правильной, возвращает нас к модели ДеГроота, поэтому далее не будем его рассматривать. Если Фробениусово число матрицы \mathbf{B} меньше единицы, то тогда она является продуктивной, мнения агентов в (1.14) сходятся и предельное распределение мнений $\mathbf{p}(\infty)$ может быть найдено из уравнения

$$\mathbf{p}(\infty) = (\mathbf{I}_n - \mathbf{X})\mathbf{W}\mathbf{p}(\infty) + \mathbf{X}\mathbf{u}.$$

Разрешая его относительно $\mathbf{p}(\infty)$, получаем:

$$\mathbf{p}(\infty) = (\mathbf{I}_n - \mathbf{B})^{-1} \mathbf{X} \mathbf{u}, \quad (1.17)$$

При этом, матрица управления $(\mathbf{I}_n - \mathbf{B})^{-1} \mathbf{X}$ является строчно-стохастической.

Динамика (1.14) позволяет оценивать влияние агентов аналогично тому, как было описано для модели ДеГроота: при условии, что мнения агентов сходятся, матрица $(\mathbf{I}_n - \mathbf{B})^{-1} \mathbf{X}$ описывает, какой вклад дают предубеждения агентов \mathbf{u} в итоговое распределение $\mathbf{p}(\infty)$ мнений. Таким образом, матрица управления $(\mathbf{I}_n - \mathbf{B})^{-1} \mathbf{X}$ может быть использована для оценки влияния агентов. Способ измерения этой величины, предложенный в [75], учитывает тот факт, что итоговое распределение мнений $\mathbf{p}(\infty)$ не обязано быть консенсусом. Для этого рассматривается, какой вклад предубеждения агентов дают в *среднее* итоговое мнение

$$\bar{p}(\infty) = \frac{1}{n} \sum_{i=1}^n p_i(\infty) = \frac{1}{n} \mathbb{1}_n^T (\mathbf{I}_n - \mathbf{B})^{-1} \mathbf{X} \mathbf{u},$$

В рамках рассматриваемого подхода, i -я компонента строки $\frac{1}{n} \mathbb{1}_n^T (\mathbf{I}_n - \mathbf{B})^{-1} \mathbf{X}$ выражает влияние агента i . При этом, представленный метод измерения влияния имеет тесную связь с известной метрикой PageRank, использующейся для ранжирования веб-страниц [65; 76; 77].

Необходимо отметить, что в отличие от большинства других моделей динамики мнений (особенно, микро-моделей), модель Фридкина-Джонсона была экспериментально подтверждена в лабораторных условиях в рамках экспериментов с малыми (четыре человека) по размеру группами [9].

Как и в случае с моделью ДеГроота, модель Фридкина-Джонсона можно обобщить по двум направлениям:

1. Мнения агентов относительно (единственного) топика описывается распределением вероятностей.
2. Рассматриваются мнения агентов относительно не одного топика, а m различных, логически независимых тем.

Оба эти обобщения формализуются с помощью уравнения

$$\mathbf{P}(t+1) = (\mathbf{I}_n - \mathbf{X}) \mathbf{W} \mathbf{P}(t) + \mathbf{X} \mathbf{U}, \quad (1.18)$$

где, как и в случае модели ДеГроота, $\mathbf{P}(t) \in \mathbb{R}^{n \times m}$ - матрица, описывающая мнения агентов в момент времени t . Ее i -я строка $\mathbf{p}_i(t) = [p_{i1}(t), \dots, p_{im}(t)]$ выражает позицию агента i . Нововведение состоит в матрице

$$\mathbf{U} = \begin{bmatrix} u_{11} & \dots & u_{1m} \\ \dots & \dots & \dots \\ u_{n1} & \dots & u_{nm} \end{bmatrix} \in \mathbb{R}^{n \times m},$$

которая представляет предубеждения агентов. На нее накладываются те же ограничения, что и на $\mathbf{P}(t)$.

Кроме того, как уже было сказано, в [60] было предложено обобщение модели Фридкина-Джонсона на случай, когда имеется несколько тем для обсуждения, и, при этом, они *логически связаны*. Для моделирования такой системы вводится матрица $\mathbf{T} \in \mathbb{R}^{m \times m}$, формализующая логические взаимосвязи между топиками. Для описания мнений агентов используется тот же подход, как и в (1.18) для случая m независимых топиков. Тогда динамика мнений агента i описывается следующим уравнением:

$$\begin{bmatrix} p_{i1}(t+1) \\ \dots \\ p_{im}(t+1) \end{bmatrix} = (1 - x_i) \mathbf{T} \left\{ w_{i1} \begin{bmatrix} p_{11}(t+1) \\ \dots \\ p_{1m}(t+1) \end{bmatrix} + \dots + w_{in} \begin{bmatrix} p_{n1}(t+1) \\ \dots \\ p_{nm}(t+1) \end{bmatrix} \right\} + x_i \begin{bmatrix} u_{i1}(t+1) \\ \dots \\ u_{im}(t+1) \end{bmatrix}.$$

Матрицу \mathbf{T} в иностранной литературе принято называть MiDS-матрицей (multi-issues dependence structure). Ее структура оказывает значительное влияние на динамику мнений. В случае, когда эта матрица является единичной, динамика системы сводится к обычной модели Фридкина-Джонсона для случая m независимых топиков. На матрицу \mathbf{T} накладываются определенные ограничения, одно из наиболее жестких – строчная стохастичность. В зависимости от требований задачи, это ограничение может быть ослаблено.

1.2.3 Протокол спорщиков

Модели ДеГроота и Фридкина-Джонсона, приведенные выше, относятся к классу *синхронно-итеративных* моделей: в каждый момент времени агенты одновременно обмениваются своими мнениями и затем, на основании полученной информации, меняют свои взгляды к следующему моменту времени. Однако, в реальных жизненных ситуациях такой формат общения маловероятен. Более реалистичной видится модель, в которой агенты взаимодействуют *парами*, а именно: в каждый момент времени два случайно выбранных агента взаимодействуют, влияя друг на друга согласно некоторому (не обязательно выпуклому) механизму формирования взглядов. Такой формат общения принято называть *протоколом спорщиков*. В [78–80] был предложен протокол спорщиков для модели Фридкина-Джонсона. Пусть заданы матрицы \mathbf{X} , \mathbf{W} и вектор \mathbf{p}_0 , при этом $\mathbf{p}(0) = \mathbf{p}_0$. Кроме того, потребуем, чтобы система (1.14) была стабильна (Фробениусово число матрицы \mathbf{B} было строго меньше единицы). Этого можно достигнуть, например, потребовав, чтобы матрица \mathbf{W} была неразложима, а матрица \mathbf{X} отличалась от нулевой. В таком случае:

$$\mathbf{p}(\infty) = (\mathbf{I}_n - \mathbf{B})^{-1} \mathbf{X} \mathbf{p}_0.$$

Агенты взаимодействуют следующим образом. В каждый момент времени t случайно выбирается ребро (j, i) графа $G[\mathbf{W}]$, которому соответствуют вершины i и j . При этом, выполнено $w_{ij} > 0$. Выбранные агенты модифицируют свои мнения согласно следующему правилу. Мнение агента i меняется на:

$$p_i(t+1) = (1 - (1 - x_i)w_{ij} - x_i w_{ij})p_i(t) + (1 - x_i)w_{ij}p_j(k) + x_i w_{ij}p_i(0),$$

в то время как мнение агента j остается неизменным.

Протокол спорщиков имеет тесную связь с динамикой модели Фридкина-Джонсона, а именно воспроизводит ее "в среднем". Обозначим

$$\tilde{\mathbf{p}}(k) = \frac{1}{k+1} \sum_{l=0}^k \mathbf{p}(l).$$

Теорема 1.2.7. [6, с. 180, теорема 23] *Выполнены следующие утверждения:*

1. Предел математического ожидания мнений агентов в момент времени t сходится к $\mathbf{p}(\infty)$:

$$\lim_{t \rightarrow \infty} \mathbb{E} [\mathbf{p}(t)] = \mathbf{p}(\infty).$$

2. Случайный процесс $\mathbf{p}(t)$ эргодичен почти наверное:

$$\mathbb{P} \left\{ \lim_{k \rightarrow \infty} \tilde{\mathbf{p}}(k) = \mathbf{p}(\infty) \right\} = 1.$$

3. Для любого $q > 0$ процесс $\mathbf{p}(t)$ L^q эргодичен:

$$\lim_{k \rightarrow \infty} \mathbb{E} [|\tilde{\mathbf{p}}(k) - \mathbf{p}(\infty)|^q] = 0.$$

1.2.4 Задача восстановления сети влияния для модели Фридкина-Джонсона

Задача восстановления сети влияния для модели Фридкина-Джонсона формулируется следующим образом [10; 81; 82]. Пусть наблюдателю известны мнения агентов рассматриваемой системы в моменты времени t_1, t_2, \dots, t_k , задаваемые серией матриц $\{\mathbf{P}(t_j)\}_{j=1}^k$, где $\mathbf{P}(t_j) \in \mathbb{R}^{n \times m}$. Эти матрицы принято называть снапшотами, указывая на то, что наблюдатель получил их, делая "снимок" системы (в реальной жизни это можно сделать с помощью проведения серии опросов). Они могут описывать мнения агентов относительно одного топика – в таком случае, как правило, они представлены в стохастической форме, соответствующей вероятностному распределению, или же относительно m разных топиков (которые могут быть как логически связаны, так и не связаны). В предположении, что динамика системы подчиняется модели Фридкина-Джонсона (1.14) (в том числе, моменты времени $\{t_j\}_{j=1}^k$, в которые были сделаны снапшоты синхронизированы с ходом времени модели), требуется восстановить матрицу влияния \mathbf{W} и матрицу \mathbf{X} . Иными словами, требуется найти все неизвестные параметры модели.

Прежде чем перейти к более подробному анализу данной задачи, обсудим ее мотивировку. Если задаться целью калибровать параметры модели модели Фридкина-Джонсона (например для ее валидации), то среди всех величин, фигурирующих в уравнении (1.14), мнения людей можно оценить, используя, например, опросы или соцсети, в то время как величины w_{ij} , задающие сеть влияния, а также коэффициенты x_i , выражающие степень приверженности людей их предубеждениям откалибровать намного сложнее. При этом, определение

этих параметров позволит во-первых, провести валидацию модели, а во-вторых, в случае адекватности модели, использовать полученные данные для прогнозирования дальнейшего развития системы.

Отдельной проблемой здесь является уже упомянутая синхронизация времени – в уравнении (1.14) мы имеем дело с абстрактным итерационным процессом, в котором ничего не сказано о масштабе изменения времени, в то время как в реальной жизни процесс изменения может развиваться с изменяющейся скоростью взаимодействий между агентами: ажиотаж перед выборами президента может сменяться затишьем после прошедших выборов и т.д.

Для решения обозначенной задачи были предложены различные подходы. Наиболее простой из них подразумевает решение следующей оптимизационной задачи. Обозначим

$$\Delta(t, \mathbf{W}, \mathbf{X}) = \mathbf{p}(t+1) - (\mathbf{I}_n - \mathbf{X})\mathbf{W}\mathbf{p}(t) - \mathbf{X}\mathbf{u}$$

и введем следующую функцию потерь

$$Q(\mathbf{W}, \mathbf{X}) = \left\| \sum_{j=1}^{k-1} \Delta(j, \mathbf{W}, \mathbf{X}) \right\|_2.$$

Тогда матрицы \mathbf{W} и \mathbf{X} могут быть получены в результате решения следующей оптимизационной проблемы

$$\min_{\mathbf{W}, \mathbf{X}} Q(\mathbf{W}, \mathbf{X}), \quad s.t. \begin{cases} \mathbf{W}\mathbb{1}_n = \mathbb{1}_n, & \mathbf{W} \geq 0, \\ \mathbf{X} = \text{diag}(x_1, \dots, x_n), & x_i \in [0, 1]. \end{cases}$$

Представленный подход имеет ряд недостатков с точки зрения практического применения. Главным из них является тот факт, что необходимо хранить все снапшоты в памяти и потеря даже одного из них может значительно повлиять на итоговый результат. Кроме того, не до конца понятно, в какие моменты времени делать снапшоты.

Альтернативный метод, предложенный в [10], учитывает обозначенные выше недостатки и предлагает рассматривать только два снапшота: начальный (p) (предполагаемый равным предубеждениям \mathbf{U}) и конечный $\mathbf{p}(\infty)$ (когда мнения стабилизируются), решая при этом задачу l_1 минимизации. Однако, существенным условием для его успешного применения является наличие информации о мнениях агентов относительно значительного числа (m одного

порядка с n) топиков, что делает данный подход трудно используемым для систем с большим числом агентов (например, для онлайн-социальных сетей).

1.2.5 Модель Краснощекова

В модели Краснощекова мнение агента i описывается вероятностью совершить некоторое действие (однако, другие интерпретации мнений, формализуемые отрезком $[0, 1]$ также уместны). Простейшая конфигурация модели, предложенная в [32], подразумевает одноитерационную процедуру обмена мнениями. Мнение агента i до общения с коллективом обозначается α_i , его мнение после общения есть p_i . Каждому агенту поставлено в соответствие число μ_i , выражающее степень его сопротивления социальному влиянию. Для агента-конформиста $\mu_i = 0$, в то время как для агента, невосприимчивого ко внешнему влиянию, выполнено $\mu_i = 1$. В результате общения с коллективом, мнение агента с $\mu_i = 1$ не меняется:

$$p_i = \alpha_i.$$

В противном случае агент меняет свое мнение, ориентируясь на мнения других агентов в соответствии со строчно-стохастической матрицей $C = [c_{ij}]$, у которой главная диагональ нулевая: $c_{ii} = 0$. Отметим, что такое ограничение имеет смысл, только если число агентов системы n больше единицы.

Определение 1.2.2. Матрицу C будем называть матрицей относительных влияний, а соответствующий ей простой ориентированный взвешенный граф $G[C]$ – сетью относительных влияний.

Для агента-конформиста выполнено

$$p_i = \sum_j c_{ij} p_j \quad (1.19)$$

В [32] рассматривался случай, когда матрица относительных влияний такова, что при $i \neq j$ выполнено $c_{ij} > 0$ то есть все агенты влияют друг на друга. Такая сеть относительных влияний моделирует небольшой коллектив, в котором

все друг с другом хорошо знакомы. Отметим, что соотношение $c_{ii} = 0$ в данном случае закономерно: агент-конформист не обращает внимания на свое мнение.

В случае агента с произвольным $\mu_i \in [0, 1]$ в силу формулы полной вероятности имеем:

$$p_i = \mu_i \alpha_i + (1 - \mu_i) \sum_{j=1}^n c_{ij} p_j. \quad (1.20)$$

В векторной формуле (1.20) принимает вид

$$\mathbf{p} = \mathbf{M}\alpha + (\mathbf{I}_n - \mathbf{M})\mathbf{C}\mathbf{p}, \quad (1.21)$$

где $\mathbf{M} = \text{diag}(\mu_1, \dots, \mu_n)$.

Замечание 1.2.1. В (1.20) и (1.21) мнение агента i зависит от **финальных** мнений других агентов. Из этого следует, что этот агент формирует свое итоговое мнение p_i **после** того, как были сформированы мнения других агентов p_j , $j \neq i$. Однако, в силу произвольности выбора агента i , это верно для каждого агента системы, что приводит к временному противоречию, которое требует дополнительного анализа. В [32] это никак не было прокомментировано и рассмотрению данной проблемы посвящен пункт 2.5.

Если известны \mathbf{M} , \mathbf{C} и α , то для нахождения итоговых мнений агентов \mathbf{p} необходимо разрешить (1.21) относительно \mathbf{p} . Решение данной задачи тесно связано с вопросом существования матрицы $(\mathbf{I}_n - (\mathbf{I}_n - \mathbf{M})\mathbf{C})^{-1}$. Кроме того, необходимо, чтобы полученные величины $p_i \in [0, 1]$. В [32] данный вопрос был подробно исследован. В силу свойств \mathbf{C} , если матрица \mathbf{M} не является нулевой, то тогда фробениусово число неразложимой матрицы $(\mathbf{I}_n - \mathbf{M})\mathbf{C}$ строго меньше единицы и она является продуктивной, что обеспечивает существование единственного решения. В противном случае решений бесконечно много, однако все они имеют чрезвычайно простую структуру.

Утверждение 1.2.1. Если матрица \mathbf{M} не является нулевой, то существует единственное решение (1.20) такое, что для любого i выполнено $p_i \in [0, 1]$. В противном случае система (1.20) имеет бесконечное число решений вида

$$\mathbf{p} = p \mathbf{1}_n. \quad (1.22)$$

Замечание 1.2.2. В (1.22) величина p может принимать любое значение из \mathbb{R} , однако, для того, чтобы оставаться в рамках вероятностных интерпретаций, необходимо дополнительно потребовать $p \in [0, 1]$.

В случае, если матрица \mathbf{M} не является нулевой, финальные мнения агентов задаются соотношением

$$\mathbf{p} = (\mathbf{I}_n - (\mathbf{I}_n - \mathbf{M})\mathbf{C})^{-1}\mathbf{M}\alpha. \quad (1.23)$$

Уравнение (1.21) описывает одну итерацию общения, в случае длительного взаимодействия агентов целесообразно рассматривать следующее рекуррентное соотношение:

$$\mathbf{p}(t+1) = \mathbf{M}\mathbf{p}(t) + (\mathbf{I}_n - \mathbf{M})\mathbf{C}\mathbf{p}(t+1). \quad (1.24)$$

Если \mathbf{M} не является нулевой, то процесс (1.24) эквивалентен

$$\mathbf{p}(t+1) = (\mathbf{I}_n - (\mathbf{I}_n - \mathbf{M})\mathbf{C})^{-1}\mathbf{M}\mathbf{p}(t) \quad (1.25)$$

с начальным условием

$$\mathbf{p}(0) = \alpha.$$

1.3 Проблема расщепления общества и феномен поляризации

В работах [33; 55] авторы Р. Абельсон и Р. Аксельрод отметили, что модели динамики мнений должны учитывать тот факт, что реальное общество никогда не находится в состоянии консенсуса: всегда есть группы людей, между которыми наблюдаются противоречия.

С точки зрения модели ДеГроота такое возможно, если сеть влияний $G[\mathbf{W}]$ не является сильно связной и в ней присутствуют по крайней мере две закрытые компоненты сильной связности. В частности, такое может быть, если в системе присутствуют агенты с $w_{ii} = 1$ – так называемые "упертые" агенты. Если рассматривать проблему расщепления общества с точки зрения модели Фридкина-Джонсона, то отсутствие консенсуса может быть объяснено приверженностью людей своим предубеждениям.

Однако, если в указанные модели интегрировать предположение о том, что матрица влияний W меняется во времени, учитывая, например, явление ограниченного доверия, то тогда отсутствие консенсуса может быть легко достигнуто в случае, если порог доверия (разница во мнениях, выше которой влияние не происходит) будет достаточно велик.

Еще одной сигнатурой нашего времени является рост уровня поляризации общественного мнения [25; 83]. Речь идет о том, что мнения индивидов отдаляются друг от друга, вызывая рост противоречий. Данный процесс имеет аналогию из теории социального неравенства: ситуация, когда бедные беднеют, а богатые становятся еще богаче. Здесь необходимо отметить, что модель ДеГроота (а, следовательно, и модель Фридкина-Джонсона) является "нерасширяющейся" (см. (1.8)). В результате, модели ДеГроота и Фридкина-Джонсона не в состоянии объяснить феномен поляризации, в отличие от, например, моделей репульсивного влияния.

Глава 2. Анализ модели Краснощекова

2.1 Случай разреженной матрицы относительных влияний

Напомним, в основополагающей работе [32] предполагалось, что матрица относительных влияний C является положительной вне главной диагонали. Это означает, что все члены коллектива знают и влияют друг на друга напрямую.

Определение 2.1.1. [65, с. с. 65] *Социальный граф — это сеть, вершины которой соответствуют людям или группам людей, а ребра — социальным связям между ними.*

Отличительной особенностью социальных графов является разреженность [65]: неформально говоря, число ребер много меньше максимально возможного их числа. К примеру, если имеется граф $G = (V, E)$, где V — множество вершин, а E — множество ребер этого графа, причем $n = |V|$ и $l = |E|$, то максимально возможное число ребер есть $\binom{n}{2}$. Если отношение $l/\binom{n}{2}$ (которое есть плотность графа ρ) достаточно мало, то граф считается разреженным.

Однако, представленное выше свойство проявляется лишь при больших n . В связи с этим, более строгое определение разреженного графа можно дать, рассматривая его эволюцию во времени, вызванную ростом числа вершин.

Определение 2.1.2. *Граф G называется разреженным, если $\rho \rightarrow 0$ при $n \rightarrow \infty$.*

Случай, рассмотренный в [32], соответствует относительно небольшому числу агентов, ведь только в случае небольшой группы людей сеть относительных влияний может быть полным графом. Однако, при моделировании более крупномасштабных систем необходимо учитывать, что сеть относительных влияний $G[C]$, которая является социальным графом, должна быть разреженной, для чего необходимо потребовать возможности равенства $c_{ij} = 0$ при некоторых $i \neq j$. Поскольку условие положительности матрицы C вне главной диагонали играло существенную роль в доказательствах утверждений в [32], учет разреженности графа $G[C]$ влечет за необходимость внесения соответствующих корректировок.

Предположение 2.1.1. *Далее предполагается, что c_{ij} может обращаться в ноль при $i \neq j$. При этом, будем считать, что для любого i существует j такой, что $c_{ij} > 0$.*

Докажем предварительно следующий вспомогательный факт.

Лемма 2.1.1. *Пусть матрица $B \in \mathbb{R}^{n \times n}$ является строчно-субстохастичной, но не строчно-стохастичной. Пусть, кроме того, все закрытые подматрицы ее нормальной формы (B') также являются строчно-субстохастичными, но не строчно-стохастичными. Тогда ее спектральный радиус (Фробениусово число) $\rho(B)$ строго меньше единицы.*

Доказательство. Поскольку матрица B является строчно-субстохастичной, но не строчно-стохастичной, следовательно, в силу [71, с. 354, теорема 6], выполнено $\rho(B) \leq 1$. Докажем, что равенство $\rho(B) = 1$ невозможно. От противного, если $\rho(B) = 1$, то тогда спектральный радиус нормальной формы матрицы B также равен единице:

$$\rho(B') = 1.$$

Но тогда спектральный радиус каждой ее закрытой подматрицы также равен единице, откуда, в силу теоремы Фробениуса-Перрона получаем противоречие. \square

Следующее утверждение представляет собой модификацию утверждения 1.2.1, которая учитывает предположение 2.1.1.

Утверждение 2.1.1. *Пусть все закрытые компоненты сильной связности сети относительных влияний $G[C]$ таковы, что каждая из них включает в себя хотя бы одного агента, не являющегося конформистом. Тогда уравнение (1.21) имеет единственное решение*

$$p = Z\alpha, \tag{2.1}$$

где матрица

$$Z = (I_n - (I_n - M)C)^{-1}M \tag{2.2}$$

является строчно-стохастической.

Доказательство. Докажем это утверждение в два этапа. Сначала покажем, что матрица \mathbf{Z} существует, а затем обоснуем, что она является строчно-стохастической.

1. Пусть, без ограничений общности, матрица \mathbf{C} уже приведена к нормальной форме. Для простоты и наглядности, будем считать, что она имеет следующий вид:

$$\mathbf{C} = \begin{bmatrix} C_{aa} & C_{ab} & C_{ac} & C_{ad} \\ 0 & C_{bb} & C_{bc} & C_{bd} \\ 0 & 0 & C_{cc} & 0 \\ 0 & 0 & 0 & C_{dd} \end{bmatrix}, \quad (2.3)$$

где квадратные подматрицы C_{aa} , C_{bb} , C_{cc} и C_{dd} неразложимы. Нижние индексы a , b , c и d выражают разбиение множества агентов на четыре группы, соответствующие компонентам сильной связности графа $G[\mathbf{C}]$. Структура матрицы \mathbf{C} , представленная в (2.3) такова, что $G[\mathbf{C}]$ содержит только две закрытых компоненты сильной связности: им соответствуют индексы c и d (подматрицы C_{cc} и C_{dd} соответственно). Декомпозиции матрицы \mathbf{C} соответствует следующая декомпозиция матрицы $\mathbf{B} = (\mathbf{I}_n - \mathbf{M})\mathbf{C}$:

$$\mathbf{B} = \begin{bmatrix} B_{aa} & B_{ab} & B_{ac} & B_{ad} \\ 0 & B_{bb} & B_{bc} & B_{bd} \\ 0 & 0 & B_{cc} & 0 \\ 0 & 0 & 0 & B_{dd} \end{bmatrix}.$$

В силу условия теоремы, в каждой из закрытых компонент сильной связности сети относительных влияний $G[\mathbf{C}]$ имеется хотя бы один агент, не являющийся конформистом. Следовательно:

1. Матрица \mathbf{B} является строчно-субстохастичной, но не строчно-стохастичной.
2. Все закрытые подматрицы нормальной формы \mathbf{B} также являются строчно-субстохастичными, но не строчно-стохастичными.

Из леммы 2.1.1 следует, что спектральный радиус матрицы \mathbf{B} меньше единицы. Отсюда следует, что существует матрица $(\mathbf{I}_n - \mathbf{B})^{-1}$, что, в свою очередь, обеспечивает существование матрицы \mathbf{Z} .

2. Теперь докажем строчную стохастичность матрицы \mathbf{Z} . Из неотрицательности матрицы \mathbf{B} следует неотрицательность \mathbf{Z} [5, с. 4, лемма 7]. Осталось

показать, что $\mathbb{1}_n$ является собственным вектором матрицы \mathbf{Z} , соответствующем собственному числу единица. Из строчной стохастичности матрицы \mathbf{C} имеем:

$$\mathbf{C}\mathbb{1}_n = \mathbb{1}_n.$$

Следовательно:

$$(\mathbf{I}_n - \mathbf{M})\mathbf{C}\mathbb{1}_n = \mathbf{B}\mathbb{1}_n = (\mathbf{I}_n - \mathbf{M})\mathbb{1}_n.$$

Отсюда получаем:

$$(\mathbf{I}_n - \mathbf{B})\mathbb{1}_n = \mathbf{M}\mathbb{1}_n.$$

Домножая слева на $(\mathbf{I}_n - \mathbf{B})^{-1}$, получаем:

$$\mathbb{1}_n = (\mathbf{I}_n - \mathbf{B})^{-1}\mathbf{M}\mathbb{1}_n = \mathbf{Z}\mathbb{1}_n,$$

ч.т.д. □

Замечание 2.1.1. *Результат, сформулированный в утверждении 2.1.1 был представлен на восьмой Международной конференции по исследованию операций в 2016 году [84]. В том же году вышла публикация [60], в которой был сформулирован более общий результат, но уже применительно к модели Фридкина-Джонсона. Таким образом, результат, сформулированный в утверждении 2.1.1 является оригинальным результатом автора настоящей диссертации.*

Замечание 2.1.2. *Тот факт, что матрица \mathbf{Z} является строчно-стохастичной, если выполнено условие утверждения 2.1.1, обеспечивает корректность вероятностных интерпретаций мнений агентов.*

В случае, когда существует закрытая компонента сильной связности, состоящая целиком из конформистов, Фробениусово число матрицы \mathbf{B} равно единице и уравнение (1.21) имеет бесконечно много решений. Опишем их структуру. Для простоты и наглядности предположим, что матрица \mathbf{C} уже приведена к нормальной форме, которая имеет следующий вид:

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{aa} & \mathbf{C}_{ab} \\ \mathbf{0} & \mathbf{C}_{bb} \end{bmatrix}, \quad (2.4)$$

где квадратные подматрицы C_{aa} и C_{bb} неразложимы. При этом, будем считать, что группа b (подматрица C^{bb}) соответствует закрытой компоненте сильной связности графа $G[C]$, полностью состоящей из конформистов. Согласованные декомпозиции матриц I_n , M и векторов α , p имеют следующий вид:

$$I_n = \begin{bmatrix} I_a & 0 \\ 0 & I_b \end{bmatrix}, M = \begin{bmatrix} M_a & 0 \\ 0 & 0 \end{bmatrix}, \alpha = \begin{bmatrix} \alpha_a \\ \alpha_b \end{bmatrix} \text{ и } p = \begin{bmatrix} p_a \\ p_b \end{bmatrix}. \quad (2.5)$$

Утверждение 2.1.2. Пусть матрицы M и C удовлетворяют (2.4), (2.5). Тогда существует матрица $Q_a = (I_a - (I_a - M_a)C_{aa})^{-1}$. При этом, уравнение (1.21) имеет бесконечно много решений вида

$$\begin{cases} p_a = \begin{bmatrix} Q_a M_a & Q_a(I_a - M_a)C_{ab} \end{bmatrix} \begin{bmatrix} \alpha_a \\ p_b \end{bmatrix}, \\ p_b = p \mathbb{1}_b. \end{cases} \quad (2.6)$$

В (2.6) p – произвольное число. При этом, прямоугольная матрица

$$\begin{bmatrix} Q_a M_a & Q_a(I_a - M_a)C_{ab} \end{bmatrix}$$

является неотрицательной и все ее строчные суммы равны единице.

Доказательство. В силу представленных выше декомпозиций, уравнение (1.21) разбивается на два:

$$\begin{cases} p_a = M_a \alpha_a + (I_a - M_a)C_{aa} p_a + (I_a - M_a)C_{ab} p_b, \\ p_b = C_{bb} p_b. \end{cases}$$

В силу строчной стохастичности матрицы C_{bb} , уравнение

$$p_b = C_{bb} p_b \quad (2.7)$$

имеет бесконечно много решений вида

$$p_b = p \mathbb{1}_b,$$

где p – произвольное число. Это следует из того, что строчно-стохастичная матрица всегда имеет собственный вектор, состоящий из единиц при собственном числе единица. Более того, из неразложимости матрицы C_{bb} в силу теоремы

Фробениуса-Перрона следует, что единица является простым корнем характеристического уравнения, а значит ей соответствует подпространство собственных векторов размерности один. Таким образом, других решений у уравнения (2.7) быть не может. В силу леммы 2.1.1, спектральный радиус матрицы $(\mathbf{I}_a - \mathbf{M}_a)\mathbf{C}_{aa}$ строго меньше единицы, что обеспечивает существование неотрицательной матрицы $(\mathbf{I}_a - (\mathbf{I}_a - \mathbf{M}_a)\mathbf{C}_{aa})^{-1}$.

Рассмотрим прямоугольную матрицу

$$\begin{bmatrix} \mathbf{Q}_a \mathbf{M}_a & \mathbf{Q}_a (\mathbf{I}_a - \mathbf{M}_a) \mathbf{C}_{ab} \end{bmatrix}, \quad (2.8)$$

существование которой следует из рассуждений выше. Заметим, что в случае, если бы в группе b присутствовал хотя бы один неконформист, то такая система удовлетворяла бы условию теоремы 2.1.1, что обеспечивало бы существование матрицы \mathbf{Z} . При этом, матрица 2.8 являлась бы подматрицей матрицы \mathbf{Z} , имея то же самое число столбцов. Это обеспечивает ее неотрицательность и равенство единице всех ее строчных сумм.

□

Пример 2.1.1. Рассмотрим систему, характеризуемую матрицами

$$\mathbf{C} = \begin{bmatrix} 0 & 1/2 & 1/4 & 1/4 \\ 1/4 & 0 & 1/2 & 1/4 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

и

$$\mathbf{M} = \begin{bmatrix} 1/2 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Условие утверждения 2.1.1 выполнено, поскольку в сети относительных влияний $G[\mathbf{C}]$ имеется одна закрытая компонента сильной связности, которой соответствует подматрица

$$\mathbf{C}_{[3:4, 3:4]} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

матрицы \mathbf{C} . При этом, в данной закрытой компоненте присутствует один неконформист: агент 4 с $\mu_4 = 1$. Следовательно, матрица \mathbf{Z} существует, является строчно-стохастичной. Ее вычисление (с округлением до второго знака после запятой) дает следующий результат:

$$\mathbf{Z} = \begin{bmatrix} 0.52 & 0.09 & 0 & 0.39 \\ 0.09 & 0.35 & 0 & 0.57 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Пример 2.1.2. Рассмотрим ту же систему, что и в примере 2.1.1, за исключением того, что матрица \mathbf{M} теперь имеет вид

$$\mathbf{M} = \begin{bmatrix} 1/2 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Данная система подходит под условие утверждения 2.1.2: в системе есть закрытая компонента сильной связности, состоящая полностью из конформистов. При этом:

$$\mathbf{M}_a = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \end{bmatrix}, \mathbf{C}_{aa} = \begin{bmatrix} 0 & 1/2 \\ 1/4 & 0 \end{bmatrix}, \mathbf{C}_{ab} = \begin{bmatrix} 1/4 & 1/4 \\ 1/2 & 1/4 \end{bmatrix} \text{ и } \mathbf{C}_{bb} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Тогда

$$\mathbf{Q}_a \mathbf{M}_a = \begin{bmatrix} 0.52 & 0.09 \\ 0.09 & 0.35 \end{bmatrix} \quad (2.9)$$

и

$$\mathbf{Q}_a (\mathbf{I}_a - \mathbf{M}_a) \mathbf{C}_{ab} = \begin{bmatrix} 0.22 & 0.17 \\ 0.37 & 0.2 \end{bmatrix}. \quad (2.10)$$

При этом, прямоугольная матрица, полученная в результате горизонтальной конкатенации матриц (2.9) и (2.10), является неотрицательной и все ее строчные суммы равны единице. В результате, решения уравнения (1.21) имеют вид:

$$\begin{cases} \mathbf{p}_a = \begin{bmatrix} 0.52 & 0.09 & 0.22 & 0.17 \\ 0.09 & 0.35 & 0.37 & 0.2 \end{bmatrix} \begin{bmatrix} \alpha_a \\ \mathbf{p}_b \end{bmatrix}, \\ \mathbf{p}_b = p \mathbb{1}_b, \end{cases}$$

где p - произвольное число.

2.2 Связь с моделями ДеГроота и Фридкина-Джонсона

Предположение 2.2.1. *Далее в этом параграфе будет предполагаться, что выполнено условие утверждения 2.1.1.*

Выпишем еще раз соотношения (1.7), (1.14) и (1.24), задающие динамики мнений в моделях ДеГроота, Фридкина-Джонсона и Краснощекова соответственно. Чтобы отличать векторы мнений в каждом из них, будем использовать соответствующие индексы:

$$\mathbf{p}_{DG}(t), \mathbf{p}_{FJ}(t), \mathbf{p}_K(t).$$

В силу предположения 2.2.1 существует матрица $\mathbf{Z} = (\mathbf{I} - (\mathbf{I} - \mathbf{M})\mathbf{C})^{-1}\mathbf{M}$, поэтому (1.24) может быть заменено на (1.25).

$$\begin{aligned} \mathbf{p}_{DG}(t+1) &= \mathbf{W}\mathbf{p}_{DG}(t), \\ \mathbf{p}_{FJ}(t+1) &= (\mathbf{I} - \mathbf{X})\mathbf{W}\mathbf{p}_{FJ}(t) + \mathbf{X}\mathbf{p}_0, \\ \mathbf{p}_K(t+1) &= (\mathbf{I} - (\mathbf{I} - \mathbf{M})\mathbf{C})^{-1}\mathbf{M}\mathbf{p}_K(t). \end{aligned} \tag{2.11}$$

Следующее утверждение демонстрирует, что одна итерация протокола модели Краснощекова может быть симитирована за счет протокола модели Фридкина-Джонсона.

Утверждение 2.2.1. *Пусть в (2.11) $\mathbf{X} := \mathbf{M}$, $\mathbf{W} := \mathbf{C}$ и $\mathbf{p}_0 := \mathbf{p}_K(t)$. Обозначим*

$$\mathbf{p}_{FJ}(\infty) = \lim_{t \rightarrow \infty} \mathbf{p}_{FJ}(t).$$

Тогда

$$\mathbf{p}_K(t+1) = \mathbf{p}_{FJ}(\infty).$$

Замечание 2.2.1. В утверждении 2.2.1 рассматривается достаточно экзотичный случай модели Фридкина-Джонсона, когда возможна ситуация $w_{ii} = 0$ и $x_i > 0$. Обычно предполагается, что $w_{ii} = 0$ тогда и только тогда, когда $x_i = 0$.

Сформулированный в утверждении 2.2.1 результат позволяет использовать свойства модели Фридкина-Джонсона при изучении модели Краснощекова. В частности, в силу теоремы 1.2.5, динамика модели Фридкина-Джонсона может быть смоделирована с помощью модели ДеГроота (с введением фиктивных агентов). Следовательно, в силу утверждения 2.2.1, протокол модели Краснощекова также может быть представлен с помощью протокола модели ДеГроота.

Теорема 2.2.1. Пусть в (2.11) матрица $\mathbf{W} \in \mathbb{R}^{2n \times 2n}$ и вектор $\mathbf{p}_{DG}(0) \in \mathbb{R}^{2n}$ таковы, что

$$\hat{\mathbf{W}} = \begin{bmatrix} (\mathbf{I} - \mathbf{M})\mathbf{C} & \mathbf{M} \\ 0 & \mathbf{I} \end{bmatrix}, \mathbf{p}_{DG}(0) = \begin{bmatrix} \mathbf{p}_K(t) \\ \mathbf{p}_K(t) \end{bmatrix}. \quad (2.12)$$

Пусть

$$\mathbf{p}_{DG}(\infty) = \lim_{t \rightarrow \infty} \mathbf{p}_{DG}(t).$$

Тогда

$$\mathbf{p}_K(t+1) = [\mathbf{p}_{DG}(\infty)]_{1:n}.$$

Этот факт, в частности, позволяет легко исследовать неотрицательность компонент матрицы $\mathbf{Z} = (\mathbf{I} - (\mathbf{I} - \mathbf{M})\mathbf{C})^{-1}\mathbf{M}$, что, в свою очередь, позволяет анализировать сходимость мнений и достижимость консенсуса в модели Краснощекова.

Замечание 2.2.2. Утверждение 2.2.1 позволяет проинтерпретировать временной масштаб в модели Краснощекова следующим образом. Динамика (1.25) описывает изменения мнений агентов на значительных временных интервалах, за которые происходит достаточно большое число итерации протокола (1.14), в ходе которых мнения агентов успевают сойтись. Далее, агенты меняют свои предубеждения, и протокол модели Фридкина-Джонсона начинается заново. Одновременно с этим начинается еще одна итерация протокола модели Краснощекова, которая заканчивается как только мнения агентов сойдутся

согласно протоколу (1.14). Необходимо уточнить, что при таком сопоставлении модель Краснощекова описывает скорее динамику **предубеждений** агентов (хотя далее, для простоты мы будем все равно использовать термин "мнение").

2.3 Сходимость мнений и достижимость консенсуса

Далее в этом параграфе считается, что выполнено предположение 2.2.1.

Проанализируем структуру строчно-стохастической матрицы $\mathbf{Z} = [z_{ij}]$, связывающей начальные и конечные мнения агентов. Более точно, нас будет интересовать вопрос, какие компоненты z_{ij} этой матрицы равны нулю и как это связано со структурой матриц \mathbf{M} и \mathbf{C} . Равенство $z_{ij} = 0$ означает, что мнение агента j не имеет влияния на мнение агента i в следующий момент времени.

Утверждение 2.3.1. При $i \neq j$ неравенство $z_{ij} > 0$ имеет место тогда и только тогда, когда в графе $G[\mathbf{C}]$ существует путь (j, \dots, i) из вершины j в вершину i такой, что $\mu_j > 0$ и для любой его вершины k , кроме, быть может, первой, выполнено $\mu_k < 1$.

Неравенство $z_{ii} > 0$ выполнено тогда и только тогда, когда $\mu_i > 0$.

Утверждение 2.3.1 позволяет исследовать сходимость мнений и достижимость консенсуса для итерационного процесса (1.25).

Теорема 2.3.1. Матрица \mathbf{Z} всегда является правильной и мнения агентов в (1.25) всегда сходятся к некоторому предельному распределению мнений $\mathbf{p}(\infty)$, которое является консенсусом тогда и только тогда, когда граф $G[\mathbf{C}]$ содержит не более одной закрытой компоненты сильной связности, в системе присутствует не более одного агента с $\mu = 1$ и если граф $G[\mathbf{C}]$ содержит одну закрытую компоненту сильной связности и в сообществе один агент с $\mu = 1$, то он должен принадлежать этой компоненте.

Пример 2.3.1. Рассмотрим систему из примера 2.1.1. Как уже было показано, в этом случае существует матрица \mathbf{Z} , которая является регулярной. Сеть относительных влияний $G[\mathbf{C}]$ содержит ровно одну закрытую компоненту сильной связности и единственный агент с $\mu = 1$ является ее членом. Для любого

начального распределения мнений такая система достигает консенсуса (см. рисунок 2.1), что подтверждается матрицей

$$\mathbf{Z}^{\infty} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

строки которой совпадают.

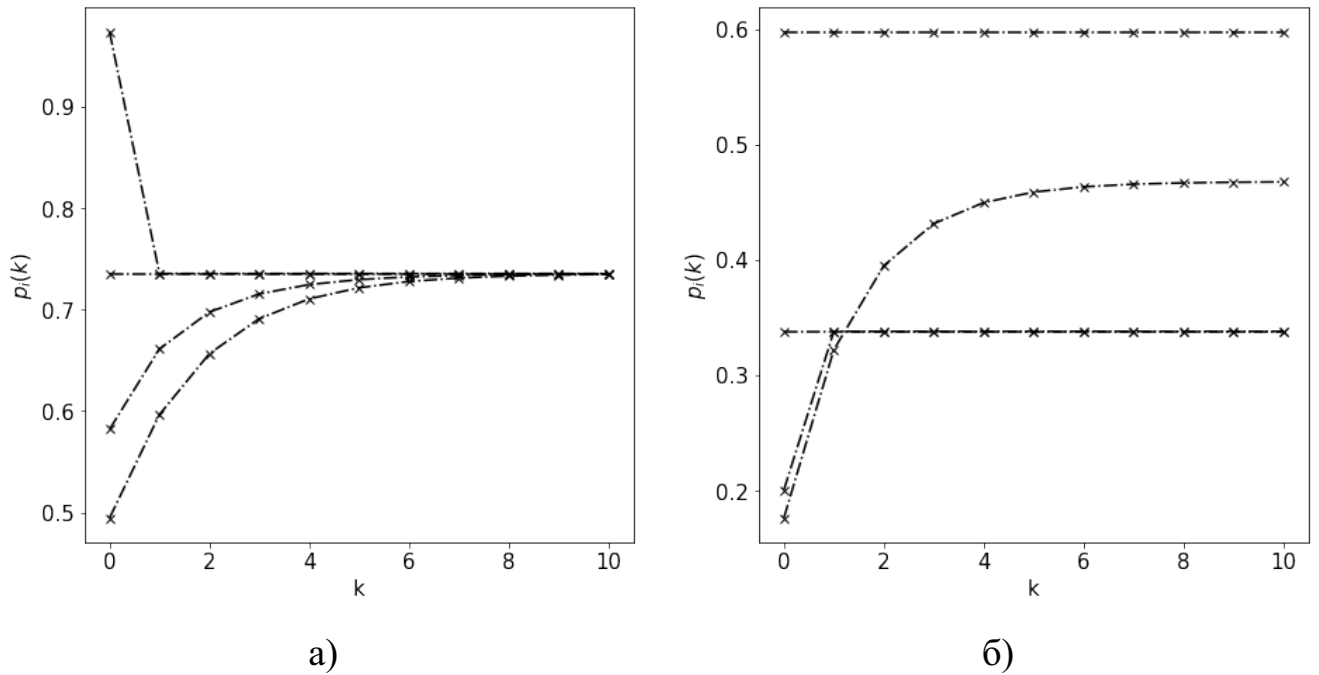


Рисунок 2.1 — (а) Динамика мнений для системы из примера 2.3.1. (б) Динамика мнений для системы из примера 2.3.2.

Пример 2.3.2. Рассмотрим систему из примера 2.1.1, однако внесем одно изменение: сделаем у агента $i = 2$ $\mu_2 = 1$. Тогда матрица \mathbf{M} будет иметь вид:

$$\mathbf{M} = \begin{bmatrix} 1/2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Как и прежде, сеть относительных влияний $G[\mathbf{C}]$ содержит ровно одну закрытую компоненту сильной связности, однако теперь в системе присутствуют два агента с $\mu = 1$: один из них является принадлежит закрытой компоненте

сильной связности, а другой нет, что приводит к отсутствию возможности достижения консенсуса (см. рисунок 2.1). Матрица

$$\mathbf{Z}^\infty = \begin{bmatrix} 0 & 0.5 & 0 & 0.5 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

строки которой не совпадают друг с другом, является тому подтверждением.

2.4 Сходимость мнений и достижимость консенсуса: случай существования изолированных конформистов, влияющих на коллектив

Рассмотрим итерационный процесс 1.24 в случае, когда не выполнено предположение 2.2.1. Для простоты будем считать, что выполнено условие утверждения 2.1.2. С учетом (2.6), динамика системы (2.4), (2.5) будет описываться следующим рекуррентным соотношением:

$$\begin{cases} \mathbf{p}_a(t+1) = \begin{bmatrix} \mathbf{Q}_a \mathbf{M}_a & \mathbf{Q}_a (\mathbf{I}_a - \mathbf{M}_a) \mathbf{C}_{ab} \end{bmatrix} \begin{bmatrix} \mathbf{p}_a(t) \\ \mathbf{p}_b(t+1) \end{bmatrix}, \\ \mathbf{p}_b(t+1) = p_{conf}(t+1) \mathbb{1}_b. \end{cases} \quad (2.13)$$

Последовательность $\{p_{conf}(t)\}_t$ в (2.13) описывает "стадные решения" конформистов. Смоделируем их, рассматривая $\{p_{conf}(t)\}_t$ как выборку из некоторого распределения с математическим ожиданием p_{conf} и дисперсией σ^2 :

$$\mathbb{E}[p_{conf}(t)] = p_{conf}, \quad \mathbb{D}[p_{conf}(t)] = \sigma^2. \quad (2.14)$$

При этом, предполагается, что мнения агентов не ограничиваются промежутком $[0, 1]$, а могут быть любым вещественным числом.

Пусть число конформистов в закрытой компоненте сильной связности равно m : $\mathbf{C}_{bb} \in \mathbb{R}^{m \times m}$. Обозначим

$$\mathbf{Z}_{aa} = \mathbf{Q}_a \mathbf{M}_a \in \mathbb{R}^{(n-m) \times n},$$

$$\mathbf{Z}_{ab} = \mathbf{Q}_a (\mathbf{I}_a - \mathbf{M}_a) \mathbf{C}_{ab} \in \mathbb{R}^{(n-m) \times m}.$$

С учетом введенных соотношений из (2.13) получаем:

$$\mathbf{p}_a(t+1) = \mathbf{Z}_{aa}\mathbf{p}_a(t) + \mathbf{Z}_{ab}\mathbf{p}_b(t+1).$$

В таком случае для $t \geq 1$:

$$\mathbf{p}_a(t) = (\mathbf{Z}_{aa})^t \mathbf{p}_a(0) + \sum_{k=1}^t (\mathbf{Z}_{aa})^{t-k} \mathbf{Z}_{ab}\mathbf{p}_b(k). \quad (2.15)$$

Пусть матрица \mathbf{Z}_{ab} является неразложимой и не является нулевой, что обеспечивает строчную субстохастичность матрицы \mathbf{Z}_{aa} . В силу того, что матрица \mathbf{Z}_{aa} неразложима и строчно-субстохастична, ее спектральный радиус меньше единицы. Следовательно

$$\lim_{t \rightarrow \infty} (\mathbf{Z}_{aa})^t = \mathbf{0}_{aa}.$$

Таким образом, для достаточно больших t верна аппроксимация

$$\mathbf{p}_a(t) \approx \sum_{k=1}^t (\mathbf{Z}_{aa})^{t-k} \mathbf{Z}_{ab}\mathbf{p}_b(k). \quad (2.16)$$

Беря математическое ожидание от обеих частей (2.15), получаем:

$$\mathbb{E} [\mathbf{p}_a(t)] = (\mathbf{Z}_{aa})^t \mathbf{p}_a(0) + p_{conf} \sum_{k=1}^t (\mathbf{Z}_{aa})^{t-k} \mathbf{Z}_{ab} \mathbb{1}_b. \quad (2.17)$$

Если матрица \mathbf{Z}_{ab} является нулевой, то тогда все тривиально, поскольку мнения агентов группы a не зависят от мнений конформистов группы b :

$$\mathbb{E} [\mathbf{p}_a(t)] = (\mathbf{Z}_{aa})^t \mathbf{p}_a(0).$$

Пусть матрица \mathbf{Z}_{ab} не является нулевой. Дополнительно потребуем, чтобы матрица \mathbf{Z}_{aa} была неразложима. В таком случае, ее спектральный радиус строго меньше единицы, что обеспечивает существование матриц $(\mathbf{I}_a - \mathbf{Z}_{aa})^{-1}$:

$$(\mathbf{I}_a - \mathbf{Z}_{aa})^{-1} = \sum_{k=0}^{\infty} (\mathbf{Z}_{aa})^k$$

и

$$(\mathbf{Z}_{aa})^{\infty} = \lim_{t \rightarrow \infty} (\mathbf{Z}_{aa})^t = \mathbf{0}_{aa}.$$

Переходя в (2.17) к пределу при $t \rightarrow \infty$, имеем:

$$\mathbb{E} [\mathbf{p}_a(\infty)] = p_{conf}(\mathbf{I}_a - \mathbf{Z}_{aa})^{-1} \mathbf{Z}_{ab}. \quad (2.18)$$

Для нахождения дисперсии случайного вектора $\mathbf{p}_a(t)$ существенным является тот факт, что элементы выборки $\{p_{conf}(t)\}_t$ независимы друг от друга. Тогда:

$$\mathbb{D} [\mathbf{p}_a(t)] = \sum_{k=1}^t \mathbb{D} [(\mathbf{Z}_{aa})^{t-k} \mathbf{Z}_{ab} \mathbf{p}_b(k)].$$

Обозначим

$$\gamma_k = (\mathbf{Z}_{aa})^{t-k} \mathbf{Z}_{ab} \mathbb{1}_b \in \mathbb{R}^{n-m}.$$

В таком случае

$$\mathbb{D} [(\mathbf{Z}_{aa})^{t-k} \mathbf{Z}_{ab} \mathbf{p}_b(k)] = \mathbb{D} [p_{conf}(t) \gamma_k] = \sigma^2 \gamma_k \circ \gamma_k.$$

С учетом этого:

$$\mathbb{D} [\mathbf{p}_a(t)] = \sigma^2 \sum_{k=1}^t \gamma_k \circ \gamma_k. \quad (2.19)$$

Пример 2.4.1. Рассмотрим систему, из утверждения 2.1.2. Детализируем матрицы \mathbf{C} и \mathbf{M} . Во первых, будем считать, что подматрицы \mathbf{C}_{aa} и \mathbf{C}_{ab} удовлетворяют

$$[\mathbf{C}_{aa}]_{ij} = 1/n - 1, \quad i \neq j,$$

$$[\mathbf{C}_{ab}]_{ij} = 1/n - 1, \quad i \neq j.$$

Кроме того, будем считать, что для $i \in [n - m]$ выполнено $\mu_i = \mu \geq 0$. Тогда, в силу симметрии, все компоненты главной диагонали матрицы \mathbf{Z}_{aa} равны между собой:

$$z_{diag}^{aa} = [\mathbf{Z}_{aa}]_{ii},$$

все остальные компоненты матрицы \mathbf{Z}_{aa} равны между собой (но необязательно равны z_{diag}^{aa}):

$$z_{off-diag}^{aa} = [\mathbf{Z}_{aa}]_{ij}, \quad i \neq j,$$

а все элементы прямоугольной матрицы \mathbf{Z}_{ab} совпадают, но в общем случае не равны z_{diag}^{aa} и $z_{off-diag}^{aa}$:

$$z^{ab} = [\mathbf{Z}_{aa}]_{ij}.$$

Тогда для любого k матрица $(\mathbf{Z}_{aa})^k$ обладает теми же свойствами, что и \mathbf{Z}_{aa} : для любых i и j

$$[(\mathbf{Z}_{aa})^k]_{ii} = [(\mathbf{Z}_{aa})^k]_{jj}$$

и для любых $i \neq j, l \neq v$

$$[(\mathbf{Z}_{aa})^k]_{ij} = [(\mathbf{Z}_{aa})^k]_{lv}.$$

В таком случае строки матрицы

$$(\mathbf{Z}_{aa})^{t-k} \mathbf{Z}_{ab}$$

совпадают. Учитывая (2.16), для достаточно больших t имеем

$$p_i(t) = p(t), \quad i \in [n - m]. \quad (2.20)$$

Из (2.20) следует, что мнения агентов неконформистов для любой наперед заданной точности начиная с некоторого момента (зависящего от этой точности), совпадают друг с другом – они достигают консенсуса, однако это состояние не является стабильным из-за влияния конформистов (см. рисунок 2.2).

Стоит отметить, что требование равенства параметров μ_i агентов существенно для достижения неустойчивого консенсуса неконформистами (см. рисунок 2.3).

Далее, для простоты будем считать, что для любого $i \in [n - m]$ выполнено

$$p_i(0) = p(0),$$

что обеспечивает (2.20) для любого t . Из (2.13) получаем:

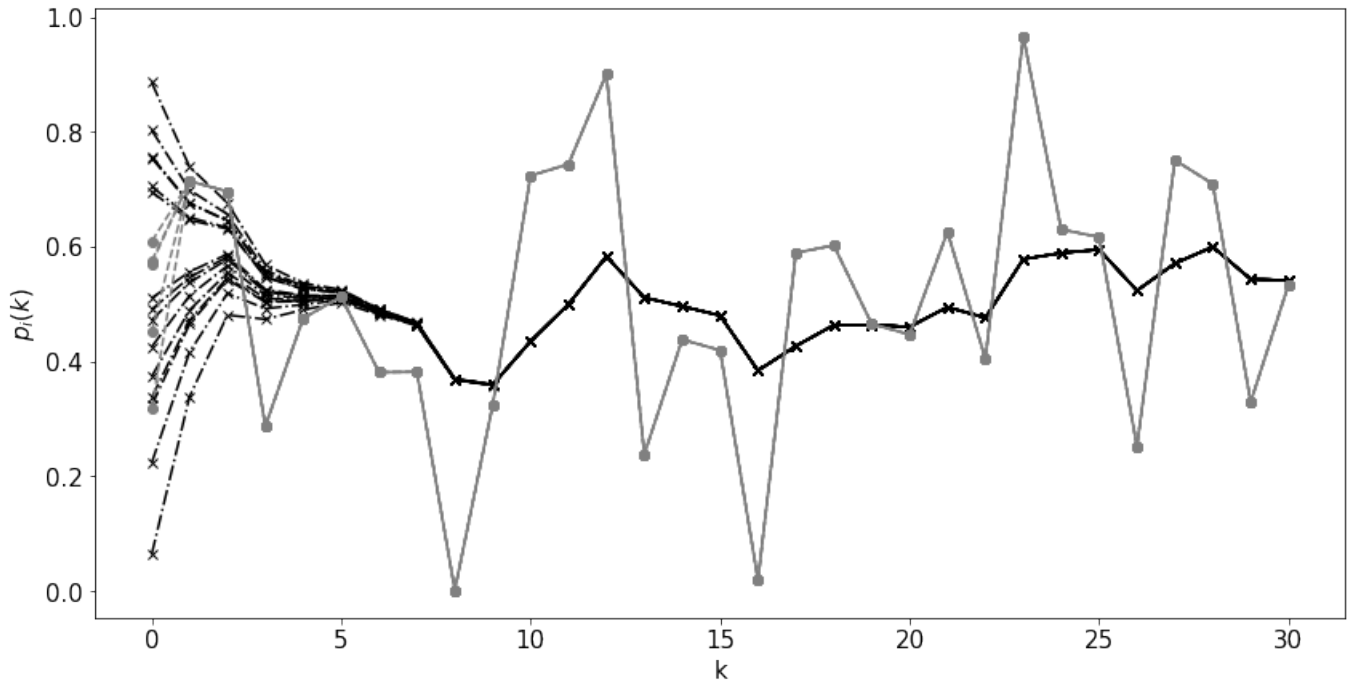


Рисунок 2.2 — Динамика мнений для $n = 20$, $m = 5$ и $\mu = 0.5$. Начальные мнения агентов откалиброваны в соответствии с нормальным распределением со средним 0.5 и стандартным отклонением 0.25. Траектории мнений агентов-конформистов представлены серым цветом, агентов-неконформистов — черным.

$$p(t+1) = \rho p(t) + (1 - \rho)p_{conf}(t+1),$$

где

$$\rho = \frac{\mu}{1 - (1 - \mu)^{\frac{n-m-1}{n-1}}} \in [0, 1].$$

При этом, $\rho = 0$ тогда и только тогда, когда $\mu = 0$. Отметим, что с ростом доли m/n конформистов параметр ρ уменьшается. Из уравнения выше для $t \geq 1$ следует:

$$p(t) = \rho^t p(0) + (1 - \rho) \sum_{k=1}^t \rho^{t-k} p_{conf}(k). \quad (2.21)$$

Найдем математическое ожидание и дисперсию случайной величины $p(t)$. Если $\rho = 1$, то тогда

$$E[p(t)] = p(0)$$

и

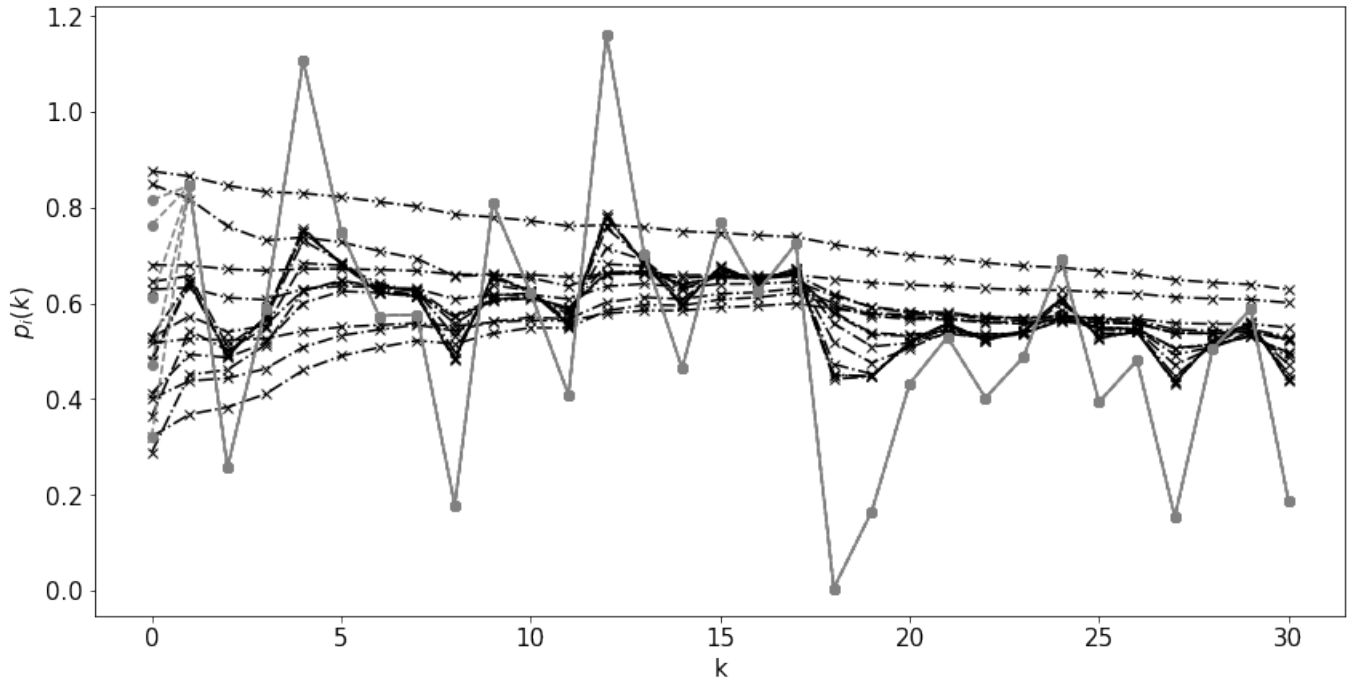


Рисунок 2.3 — Динамика мнений для той же системы, что представлена на рисунке 2.2 за исключением того, что коэффициенты μ_i агентов-неконформистов калибруются как случайная величина, равномерно распределенная на отрезке $[0, 1]$.

$$D[p(t)] = 0.$$

Далее будем предполагать, если не оговорено обратное, что $\rho < 1$. Беря математическое ожидание от обеих частей уравнения (2.21), получаем:

$$\begin{aligned} E[p(t)] &= \rho^t p(0) + (1 - \rho) p_{conf} \sum_{k=1}^t \rho^{t-k} = \\ &= \rho^t p(0) + (1 - \rho) p_{conf} \frac{1 - \rho^t}{1 - \rho} = \\ &= \rho^t p(0) + (1 - \rho^t) p_{conf}. \end{aligned} \quad (2.22)$$

При $t \rightarrow \infty$ имеем:

$$E[p(\infty)] = p_{conf}.$$

Аналогично для дисперсии:

$$\begin{aligned}
D[p(t)] &= D[\rho^t p(0) + (1 - \rho) \sum_{k=1}^t \rho^{t-k} p_{conf}(k)] = \\
&= D[(1 - \rho) \sum_{k=1}^t \rho^{t-k} p_{conf}(k)] = \\
&= (1 - \rho) \sum_{k=1}^t \rho^{t-k} D[p_{conf}(k)] = \\
&= (1 - \rho) \sigma^2 \sum_{k=1}^t \rho^{t-k} = \\
&= (1 - \rho)^2 \sigma^2 \frac{1 - \rho^{2t}}{1 - \rho^2} = \\
&= (1 - \rho^{2t}) \sigma^2 \frac{1 - \rho}{1 + \rho}.
\end{aligned} \tag{2.23}$$

При $t \rightarrow \infty$:

$$D[p(\infty)] = \sigma^2 \frac{1 - \rho}{1 + \rho}.$$

Заметим, что с ростом ρ дисперсия величины $p(t)$ уменьшается. Это следует из того, что

$$\begin{aligned}
&\left[\frac{D[p(t)]}{\sigma^2} \right]'_{\rho} = \\
&= \left[(1 - \rho^{2t}) \frac{1 - \rho}{1 + \rho} \right]'_{\rho} = \\
&= (1 - \rho^{2t}) \left[\frac{1 - \rho}{1 + \rho} \right]'_{\rho} - 2t \rho^{2t-1} \frac{1 - \rho}{1 + \rho} = \\
&= -2t \rho^{2t-1} \frac{1 - \rho}{1 + \rho} - 2 \frac{1 - \rho^{2t}}{(1 + \rho)^2} < 0
\end{aligned}$$

при $\rho \in [0, 1)$.

Имеет место следующее равенство

$$\lim_{\rho \rightarrow 1-0} [D[p(t)]]'_{\rho} = 0.$$

При $\rho = 0$ дисперсия максимальна и равна

$$D[p(t)] \Big|_{\rho=0} = \sigma^2,$$

а при $\rho = 1$ она достигает минимально возможного значения

$$D[p(t)] \Big|_{\rho=1} = 0.$$

Обозначим A_t событие, состоящее в том, что такая система достигает консенсуса в момент времени t (который, однако, не будет ее стационарным состоянием). Тогда:

$$P\{A_t\} = P\{p(t) = p_{conf}(t)\}. \tag{2.24}$$

Отметим, что если $\mu = 0$, то тогда все тривиально: в каждый момент времени t будет достигаться новый консенсус: модель общества конформистов, которым руководят также конформисты. Далее будем считать, что $\mu > 0$.

В случае непрерывной случайной величины $p_{conf}(t)$ вероятность события (2.24) равна нулю и имеет смысл говорить не о вероятности консенсуса, а о возможности события, что в момент времени t мнения конформистов и всех остальных будут отличаться не более чем на некоторое $\varepsilon > 0$:

$$P \{A_t^\varepsilon\} = P \{|p(t) - p_{conf}(t)| < \varepsilon\}.$$

Рассмотрим случайную величину

$$\xi(t) := p(t) - p_{conf}(t),$$

равную разности случайных величин $p(t)$ и $p_{conf}(t)$. В силу (2.14) и (2.23), ее математическое ожидание есть

$$E[\xi(t)] = E[p(t)] - E[p_{conf}(t)] = \rho^t(p(0) - p_{conf}).$$

При $t \rightarrow \infty$ получаем:

$$E[\xi(\infty)] = 0. \quad (2.25)$$

Дисперсия случайной величины $\xi(t)$ описывается соотношением

$$D[\xi(t)] = D[p(t)] + D[p_{conf}(t)] - 2\text{Cov}(p(t), p_{conf}(t)).$$

При этом, учитывая (2.14) и (2.23), можем для $\text{Cov}(p(t), p_{conf}(t))$ выписать следующее соотношение:

$$\begin{aligned} \text{Cov}(p(t), p_{conf}(t)) &= \text{Cov}(\rho^t p(0) + (1 - \rho) \sum_{k=1}^t \rho^{t-k} p_{conf}(k), p_{conf}(t)) = \\ &= \text{Cov}((1 - \rho) \sum_{k=1}^t \rho^{t-k} p_{conf}(k), p_{conf}(t)) = \\ &= (1 - \rho) \sum_{k=1}^t \rho^{t-k} \text{Cov}(p_{conf}(k), p_{conf}(t)) = \\ &= (1 - \rho) \text{Cov}(p_{conf}(t), p_{conf}(t)) = \\ &= (1 - \rho) \sigma^2. \end{aligned}$$

Отсюда следует

$$D [\xi(t)] = \sigma^2 \frac{(1 - \rho^{2t})(2\rho^2)(1 - \rho) + (2\rho - 1)(1 + \rho)}{1 + \rho}.$$

и

$$D [\xi(\infty)] = \sigma^2 \frac{2\rho^2}{1 + \rho}. \quad (2.26)$$

В силу неравенства Чебышева, из (2.25) и (2.26) получаем:

$$P \{ |\xi(\infty) - E [\xi(\infty)]| \geq \varepsilon \} = P \{ |\xi(\infty)| \geq \varepsilon \} = 1 - P \{ A_\infty^\varepsilon \} \leq \frac{2\sigma^2\rho^2}{(1 + \rho)\varepsilon^2}.$$

Тогда справедлива следующая оценка (см. рисунок 2.4):

$$P \{ A_\infty^\varepsilon \} \geq 1 - \frac{2\sigma^2\rho^2}{(1 + \rho)\varepsilon^2}. \quad (2.27)$$

2.5 Разрешение временного парадокса в модели Краснощекова

Рассмотрим подробнее временной парадокс, упомянутый в параграфе 1.2.5 (см. замечание 1.2.1). Из утверждения 2.2.1 и теоремы 1.2.7 следует, что если выполнено предположение 2.2.1, то одна итерация взаимодействия агентов в модели Краснощекова может быть симитирована "в среднем" за счет протокола спорщиков с соответствующими параметрами. При этом, мнения спорщиков "в среднем" будут сходиться к вектору $(I - (I - M)C)^{-1}M\alpha$, описывающему мнения агентов после общения в рамках (1.23). Это обстоятельство может служить разрешением временного парадокса, которое хорошо согласуется с замечанием 2.2.2: поскольку одна итерация протокола (1.20) соответствует достаточно продолжительному промежутку времени, должен быть учтен тот факт, что взаимодействия агентов происходят несинхронно и стохастично.

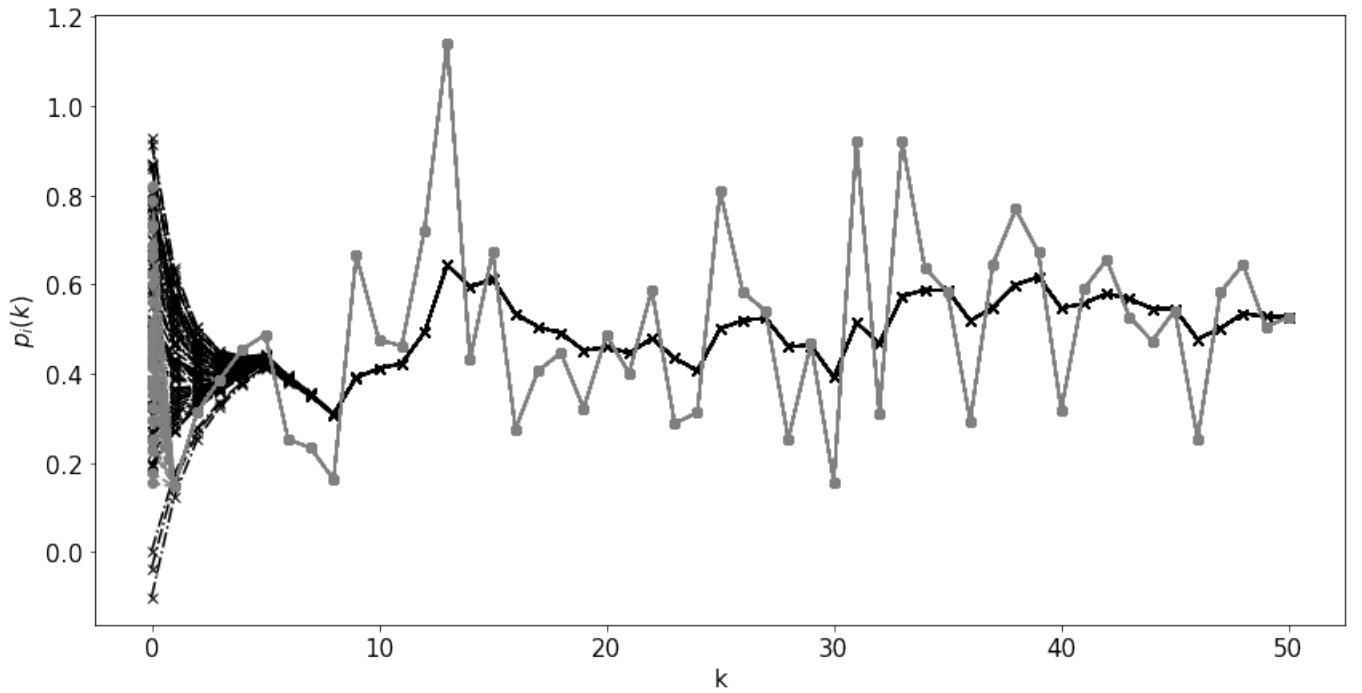


Рисунок 2.4 — Динамика мнений для системы из примера 2.4.1. Параметры системы калиброваны следующим образом. Число агентов $n = 100$. Число конформистов $m = 30$. Коэффициент сопротивления социальному влиянию агентов-неконформистов $\mu = 1/2$. Мнения агентов в начальный момент времени взяты из нормального распределения со средним 0.5 и дисперсией $\sigma^2 = 0.04$. Мнение конформистов $p_{conf}(t)$ есть случайная величина, подчиняющаяся нормальному распределению со средним 0.5 и дисперсией $\sigma^2 = 0.04$. Рассматривается процесс из 50 - ти итераций. Траектории мнений агентов-конформистов показаны серым, остальных агентов - черным. Для такой системы $\rho = 0.7$. К моменту седьмой итерации агенты-неконформисты достигают консенсуса. Если взять $\varepsilon = 0.3$, то для $t \in \{25, 26 \dots, 50\}$ доля итераций, в рамках которых $|p(t) - p_{conf}(t)| < \varepsilon$ составляет 0.88. При этом, оценка снизу (2.27) равняется 0.77.

2.6 Модель Краснощекова: расщепление общества и поляризация мнений

Результаты, представленные в данной главе свидетельствуют о том, что в случае, когда в рассматриваемой системе нет агентов с $\mu \in \{0, 1\}$ и сеть относительных влияний $G[C]$ является сильно связным графом, динамика мнений, подчиняющаяся модели Краснощекова неизбежно ведет к консенсусу, что не может объяснить как явление расщепление общества, так и феномен поляризации.

При этом, наличие хотя бы двух закрытых компонент сильной связности графа $G[C]$ или хотя бы двух агентов с $\mu = 1$ обеспечит кластеризацию мнений.

Что касается явления поляризации общества, то присутствие в системе двух закрытых групп, состоящий из конформистов может привести к временной поляризации общества, однако, подобная система видится слишком "экзотичной".

Глава 3. Анализ динамики мнений пользователей социальной сети ВКонтакте

3.1 Краткий обзор методологии

В данной главе приведены результаты анализа реальных данных о динамике мнений пользователей ОСС сети ВКонтакте с точки зрения моделей Краснощекова и ДеГроота. Необходимо отметить, что применение модели Краснощекова стало возможным после ее обобщения на случай разреженной матрицы относительных влияний.

Предлагаемый подход заключается в том, чтобы использовать информацию с аккаунтов пользователей для калибровки параметров моделей. Функционал ВКонтакте подразумевает два типа связей между аккаунтами: (1) подписка и (2) дружба. Первый тип связи является односторонним, в то время как второй – двунаправленный и соответствует более высокому уровню доверия между пользователями: два пользователя становятся друзьями в ВКонтакте, если они подписаны друг на друга. В ВКонтакте существуют четыре типа аккаунтов:

1. Аккаунты пользователей;
2. Аккаунты публичных страниц;
3. Аккаунты групп;
4. Аккаунты событий.

Публичные страницы, на которые подписан пользователь, а также другие пользователи у кого более 1000 подписчиков, на которых он подписан (не его друзья) формируют блок "Интересные страницы", присутствующее на странице данного пользователя. Интересные страницы (далее – ИС) являются одними из главных источников информации в ВКонтакте. В связи с этим ИС, на которые подписан пользователь, были использованы для оценки его текущего мнения (см. параграф 3.2), в то время как граф знакомств, фиксирующий дружеские связи ВКонтакте – для оценки сети влияния. В качестве топика была выбрана политическая тематика. В связи с этим, для оценки также использовались манифестированные пользователем политические взгляды (см. далее).

Предопсылки для такого подхода следующие. Во-первых, в соответствие с уже имеющимися результатами в области предсказания политических взглядов

по данным из ОСС, ИС, на которые подписан пользователь тесно связаны с его политическими предпочтениями [18–20; 25]. Кроме того, между ИС и синтетическими агентами из теоремы 1.2.5, которые отвечают за предубеждения агента в модели Фридкина-Джонсона, можно провести аналогию: в обоих случаях речь идет о постоянном одностороннем влиянии, оказываемом на индивида. В связи с этим, учитывая замечание 2.2.2, применение именно модели Краснощекова для анализа динамики такой системы видится особенно перспективным.

Что касается графа дружеских связей ВКонтакте, нельзя утверждать что его структура совпадает со структурой сети (относительных) влияний (как минимум, потому что граф дружбы описывает только открытые настройками приватности ВКонтакте связи, упуская другие типы общения, в том числе вне ВКонтакте). Кроме того, граф знакомств является невзвешенным графом, который необходимо перед использованием еще насытить весами ребер. Тем не менее, его свойства тесно связаны с сетью влияний, поэтому в первом приближении можно воспользоваться им.

3.2 Предсказательная модель для определения политических взглядов пользователей ВКонтакте

В [85] была предложена предсказательная модель для определения политических взглядов пользователей онлайн-социальной сети ВКонтакте (далее – Предсказательная Модель), в основе которой лежит методология обучения с учителем. В ходе решения задачи 4-х классовой классификации, каждому пользователю i она ставит в соответствие стохастический вектор

$$\tilde{\mathbf{p}}_i = [\tilde{p}_i^1, \tilde{p}_i^2, \tilde{p}_i^3, \tilde{p}_i^4]^T, \quad (3.1)$$

компоненты которого отвечают за приверженность пользователя следующим политическим фигурам (фамилии перечислены в алфавитном порядке): П.Н. Грудинин, В.В. Жириновский, А.А. Навальный и В.В. Путин. Например, вектор

$$\tilde{\mathbf{p}}_i = [0, 0.25, 0, 0.75]^T$$

свидетельствует о том, что пользователь i является убежденным сторонником В.В. Путина, имеющим, при этом, слабо выраженные симпатии к В.В. Жириновскому. С точки зрения задачи классификации, такой пользователь будет классифицирован как представитель электората В.В. Путина.

Далее в этом параграфе будет представлено подробное описание Предсказательной модели, результаты ее тестирования и апробации.

Соответствующие вспомогательные материалы (в том числе разработанные комплексы программ) находятся в открытом репозитории [86]. Необходимо отметить, что в [85; 86] было предложено две предсказательные модели (там они обозначаются как ЛГ- и НС-модели). Оригинальным результатом автора настоящей диссертации является ЛГ-модель, а также алгоритмы и комплексы программ, связанные с выгрузкой размеченных данных, их разведочным анализом, кодировкой, настройкой гиперпараметров Предсказательной Модели, ее апробацией и анализом полученных результатов. Единственный результат, представленный в [85; 86] и не являющийся оригинальным результатом автора настоящей диссертации – НС-модель.

Вся работа была проведена на 64-битной машине, имеющей следующие характеристики: 24GB RAM, GPU Nvidia GeForce 1070, CPU Intel Xeon 3.07GHz. Окружение: Jupiter Notebook 5.4.0.

3.2.1 Датасеты

Были использованы два датасета. Первый датасет (Д1) состоял из "размеченных" пользователей: тех чьи политические предпочтения считаются известными заранее; данный датасет необходим для обучения и тестирования моделей машинного обучения, настройки гиперпараметров, в результате чего выбирается наилучшая (далее - Предсказательная Модель).

Д1 был создан в результате выгрузки данных, имевшей место 20 декабря 2017 года. Для конструирования Д1 был использован основополагающий результат из теории селективной экспозиции [87; 88]: индивид стремится избегать информации, которая конфликтует с его собственным воззрением, отдавая предпочтение тому, что согласуется с его взглядами. В связи с этим, предполагалось, что пользователь, подписанный на политический аккаунт, поддерживает соот-

ветствующую политическую фигуру. К примеру, пользователь, подписанный на официальный аккаунт В.В. Жириновского, является его сторонником. Необходимо отметить, что данное предположение неприменимо к каждому политически ангажированному пользователю: возможна ситуация, при которой пользователь подписан на аккаунт политика, деятельность которого пользователь не одобряет. Однако, суть данного предположения состоит в том, что **для большого числа** пользователей, оно выполняется с достаточно высокой частотой.

Рассматривался следующий список политических фигур (фамилии в алфавитном порядке): П.Н. Грудинин, В.В. Жириновский, А.А. Навальный и В.В. Путин. Соответствующее пространство меток классов \hat{P} состоит из четырех элементов. Далее, для простоты, будут рассматриваться следующие обозначения: класс 1 — сторонники П.Н. Грудинина, класс 2 — сторонники В.В. Жириновского, класс 3 — сторонники А.А. Навального, класс 4 — сторонники В.В. Путина:

$$\hat{P} = \{1, 2, 3, 4\}. \quad (3.2)$$

Данный набор классов практически полностью покрывает множество тех кандидатов в президенты на выборах 2018 года (участвовавших или не допущенных к ним), чьи аккаунты-представители в ВКонтakte имели достаточное число подписчиков на момент скачивания. Существенным ограничением такого подхода к формализации пространства мнений является отсутствие в пространстве политических взглядов (3.2) тех мнений, которые соответствуют политически неангажированным пользователям или индивидам, чьи политические предпочтения нейтральны.

Определение 3.2.1. *Маркерными аккаунтами* будем называть аккаунты, чьи подписчики были использованы для наполнения Д1

Состав Д1 был сформирован следующим образом. На момент скачивания у В.В. Жириновского и А.А. Навального были официальные аккаунты в ВКонтakte, их URL-адреса: <https://vk.com/zhirinovskiyv> и <https://vk.com/navalny> соответственно. С каждого из них были выбраны случайно по 100000 подписчиков и с этих аккаунтов была выгружена вся необходимая информация. При этом, удаленные/заблокированные/закрытые настройками приватности аккаунты, а также те пользователи, которые посещал ВКонтakte последний раз более чем за месяц до момента выгрузки, были исключены из выборок, поэтому их реальные размеры немного меньше, чем 100000 (см. таблицу 1). У П.Н.

Грудинина также был аккаунт <https://vk.com/id140204185>. Однако он не являлся официальным, а также имел всего около 17 тысяч подписчиков (для сравнения у В.В. Жириновского их около 446 тысяч), которые были выкачаны как представители класса 1. Вследствие этого, было принято решение наряду с подписчиками П.Н. Грудинина учитывать подписчиков официального аккаунта Г.А. Зюганова https://vk.com/gennadiy_zyuganov (порядка 114 тысяч подписчиков), рассматривая составную политическую фигуру “Грудинин + Зюганов” (или, просто, партию КПРФ). Далее, для удобства эти пользователи также будут называться представителями класса 1. В итоге, 83 тысячи подписчиков Г.А. Зюганова были собраны в Д1 для наполнения класса 1. У Президента В.В. Путина не было официального аккаунта в Вконтакте, однако имелось множество достаточно популярных онлайн-сообществ, которые освещали его деятельность. Из них были отобраны вручную три наиболее популярные и активные (в которых публикуются посты): https://vk.com/vladimir_vladimirovichp, https://vk.com/putin_z и https://vk.com/moy_putin. С каждого из них случайным образом было выбрано по 50 тысяч подписчиков, которые в дальнейшем рассматривались как представители класса 4.

Среди пользователей Д1 не было таких, которые принадлежали одновременно двум классам. При этом, среди множества представителей класса 1 имелось 736 повторений, а среди представителей класса 4 — 700, что связано с тем, что данные два класса были сформированы с помощью двух и трех маркерных аккаунтов соответственно.

Таблица 1 — Распределение пользователей Д1 по классам

Класс	Численность класса
Класс 1	92632
Класс 2	92144
Класс 3	93951
Класс 4	138551
Всего пользователей	417278

Второй датасет (Д2) был сконструирован из 22288537 пользователей, удовлетворяющих следующим условиям:

- Аккаунт пользователя не удален, не забанен, не закрыт настройками приватности;
- Пользователь из Российской Федерации;
- Пользователь старше 18-ти лет;
- Пользователь посещал ВКонтакте хотя бы раз за последние 28 дней (на момент загрузки).

Данный датасет использовался для апробации Предсказательной Модели. Д2 был создан в результате загрузки данных, имевшей место 20 февраля 2018 года, поэтому должен отражать предвыборные настроения рассматриваемой выборки. Данные по Д2 были предоставлены лабораторией Наук о больших данных и проблемах общества, ТГУ, Томск, Российская Федерация. После применения Предсказательной Модели к Д2 была получена оценка распределения симпатий рассматриваемых пользователей в преддверии выборов.

3.2.2 Признаки

Признаки – это характеристики аккаунтов, которые использовались как независимые переменные при решении задачи классификации. Будучи закодированными, они сформировали матрицу объекты-признаки. Одним из ключевых шагов, необходимых для успешного решения задачи классификации, является выбор наиболее информативных независимых переменных. Использовались следующие признаки:

1. Подписки на ИС. В соответствии с теорией селективной экспозиции [87–89], подписки пользователя на ИС потенциально являются чрезвычайно информативным типом признаков. В частности, в смежных работах [18; 20] подход, основанный на использовании только подписок пользователя, был успешно применен для решения аналогичной задачи.
2. Выбранные пользователем политические предпочтения из списка, предлагаемого платформой (поле "политические предпочтения" – его значения имеют принципиально иной характер, чем те, на основании которых сконструировано пространство мнений (3.2)). Чтобы отличать данный признак от зависимой переменной, далее для его обозначения будет использоваться сокращение ПП.

Кроме того, для проведения разведочного анализа у пользователей были выгружены пол и возраст.

3.2.3 Выгрузка данных из Вконтакте

Для выгрузки данных Д1 использовалась библиотека vk, позволяющую работать с методами VK API.

Пол, возраст и ПП пользователей выгружались с помощью процедуры users.get(). Информация о поле пользователей была выгружена в следующем виде:

- 0 – не указан.
- 1 – женский.
- 2 – мужской.

Информация касательно возраста пользователей была выгружена приведенным ниже образом:

- -1 – не указан.
- Иначе - актуальный возраст в годах.

Данные, связанные с ПП пользователей скачивались согласно следующим обозначениям:

- -1 – не указаны.
- 1 – “Индифферентные”.
- 2 – “Коммунистические”.
- 3 – “Социалистические”.
- 4 – “Умеренные”.
- 5 – “Либеральные”.
- 6 – “Консервативные”.
- 7 – “Монархические”.
- 8 – “Ультраконсервативные”.
- 9 – “Либертарианские”.

У пользователей есть возможность выбрать другие ПП, которых нет в представленном выше списке, однако пользователей в Д1, поступившим таким образом, всего пять, поэтому предполагалось, что они не выбрали свои политические взгляды.

Для загрузки ИС использовалась процедура `users.getSubscriptions()`. Данный функционал загружает ИС пользователя в том порядке, в котором они расположены на его странице, а он, в свою очередь, связан с активностью пользователя по отношению к данным ИС. Имеет место следующая закономерность: чем выше позиция аккаунта в списке ИС, тем выше уровень активности пользователя по отношению к данному аккаунту. Однако, возможны исключения (согласно https://vk.com/page-2158488_49931879). Учитывая это, а также ограниченное внимание пользователя [90; 91] и алгоритмы ранжирования контента, которые использует ВКонтакте (<https://vk.com/@stasfitn-novyi-algoritm-ranzhirovaniya-postov-vkontakte>), для каждого аккаунта Д1 были выгружены только первые 200 ИС (если такие имелись).

3.2.4 Разведочный анализ данных

На рисунке 3.1 представлена половозрастная пирамида пользователей Д1. На рисунке 3.2 изображено распределение пользователей относительно политических предпочтений. Гистограммы, описывающие подписки пользователей на ИС показаны на рисунке 3.3. Видно, что возрастная структура Д1 смещена в сторону более молодых представителей. Также налицо дисбаланс в сторону мужского населения и пользователей, имеющих умеренные политические взгляды. Не стоит удивляться наличию пользователей с возрастом более, чем 90 и даже 100 лет, ведь пользователь не обязан заполнять свой настоящий возраст.

Сравнение половозрастного состава пользователей Д2 с половозрастным составом населения Российской Федерации представлено на рисунке 3.4. Налицо преобладание молодых пользователей, а также сбалансированность по полу (в отличие от Д1).

Рассмотрим, насколько часто представлены рассматриваемые признаки, а также пол и возраст пользователей на их аккаунтах.

Определение 3.2.2. *Надежностью признака для данного датасета будем называть относительную частоту его присутствия на страницах пользователей.*

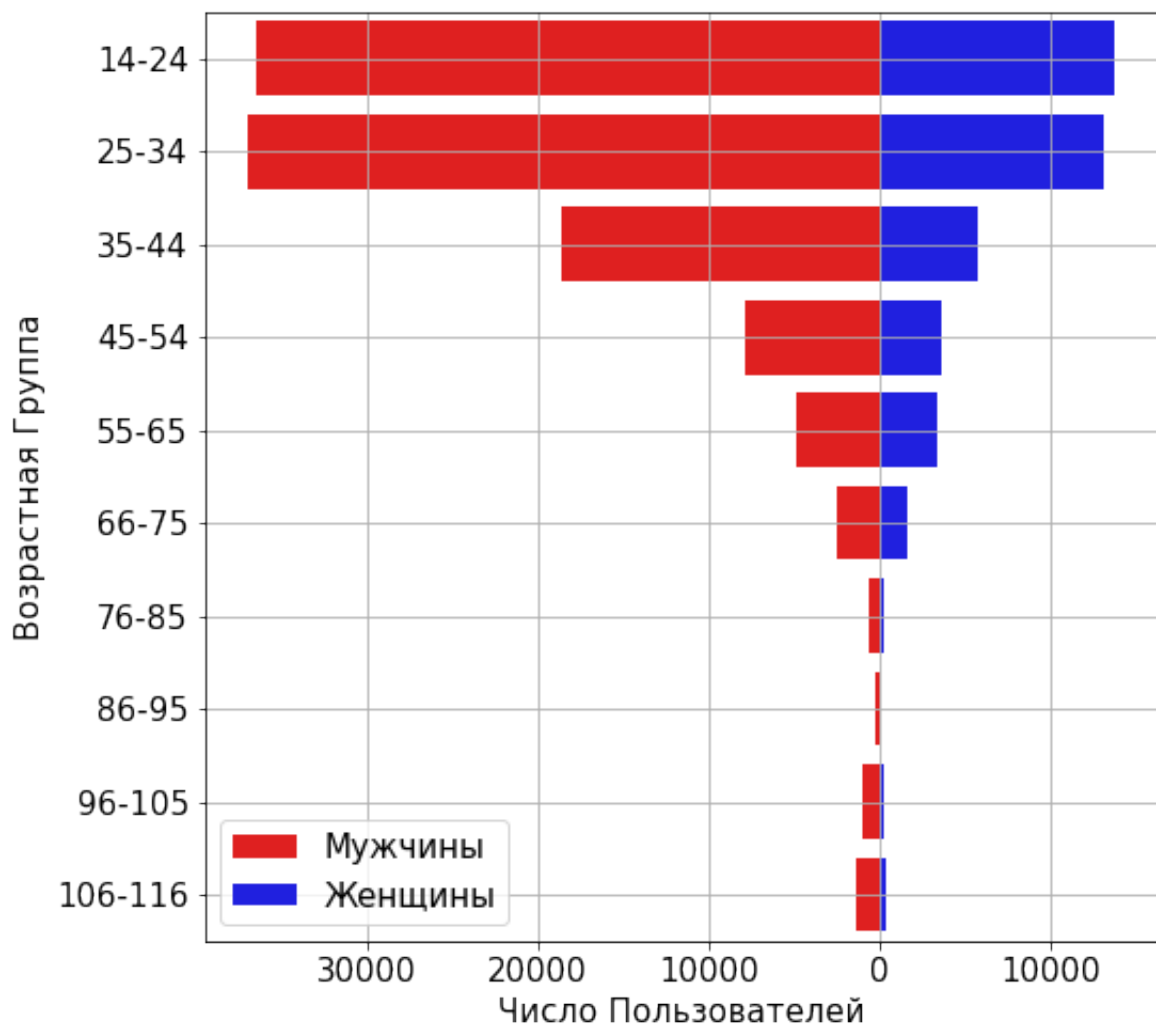


Рисунок 3.1 — Половозрастная пирамида пользователей Д1.

Отсутствие признака на странице может быть вызвано тем, что пользователь не подписывался ИС, или же при регистрации и в дальнейшем не счел нужным заполнить поле "политические предпочтения". В силу того, что пользователи, закрывшие свои аккаунты настройками приватности, не рассматривались – такие аккаунты отсеивались при выгрузке – других возможных причин отсутствия признака на странице пользователя нет. В рамках Д1 у 22.1 процентов пользователей нет ни одной подписки на ИС – надежность признака ИС равна 77,9 процентов. В то время как ИС пользователей является относительно надежными признаками, признак ПП, хоть и видится чрезвычайно важным в контексте рассматриваемой задачи, намного реже заполняется пользователями (см. таблицу 2) и, следовательно, не является надежным источником для построения прогноза. При этом, возраст и пол указаны у 36.7 и 98.2 процентов пользователей Д1 соответственно.

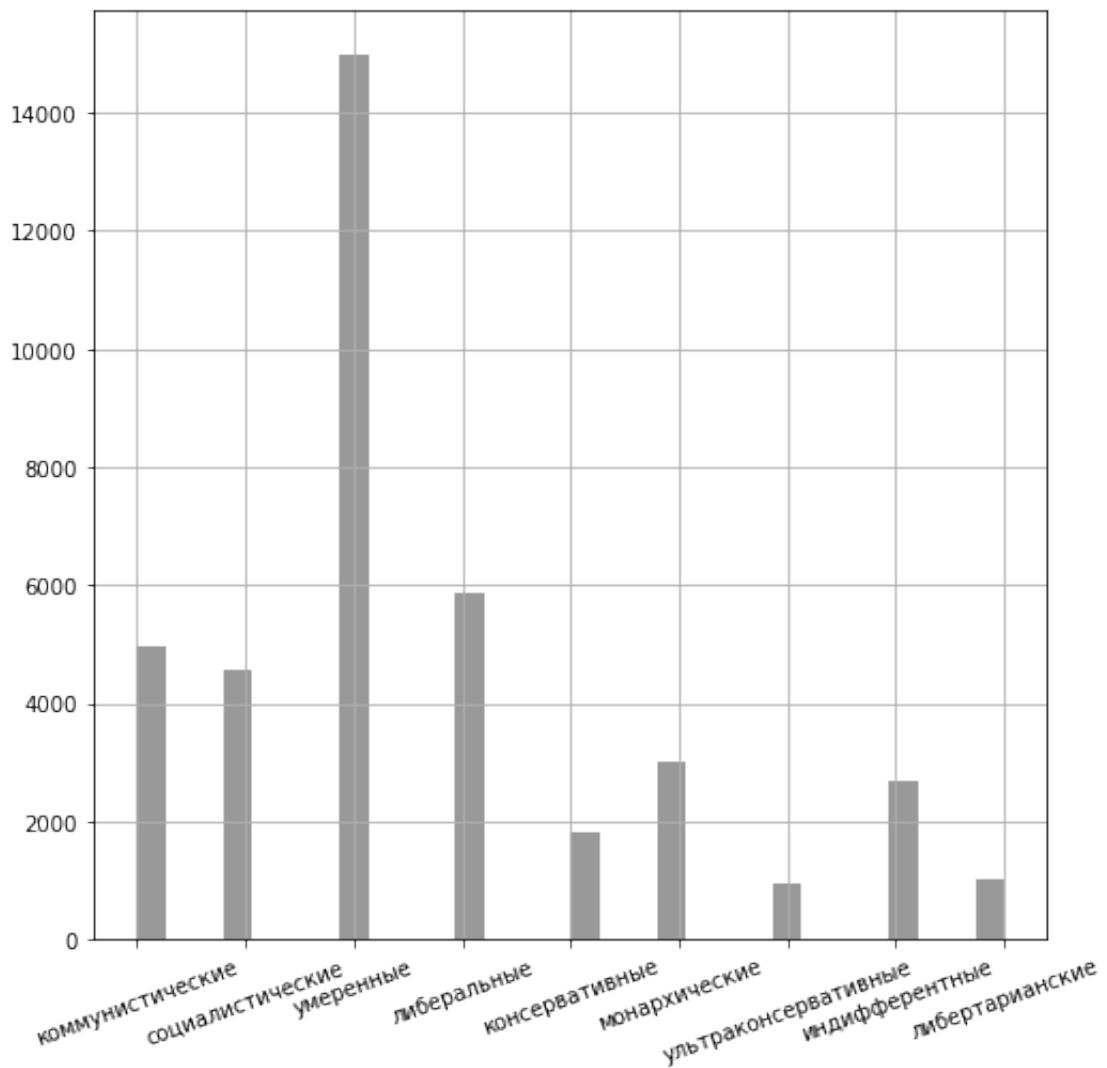


Рисунок 3.2 — ПП пользователей Д1.

Таблица 2 — Надежность признаков для пользователей Д1

Признак	Надежность
ИС	0.779
ПП	0.195

3.2.5 Кодировка

Для обработки признаков использовалось one-hot кодирование [92—94], при котором каждый признак формализуется с помощью нулей и единиц. Рассмотрим признак, имеющий k возможных значений (сюда также значение,

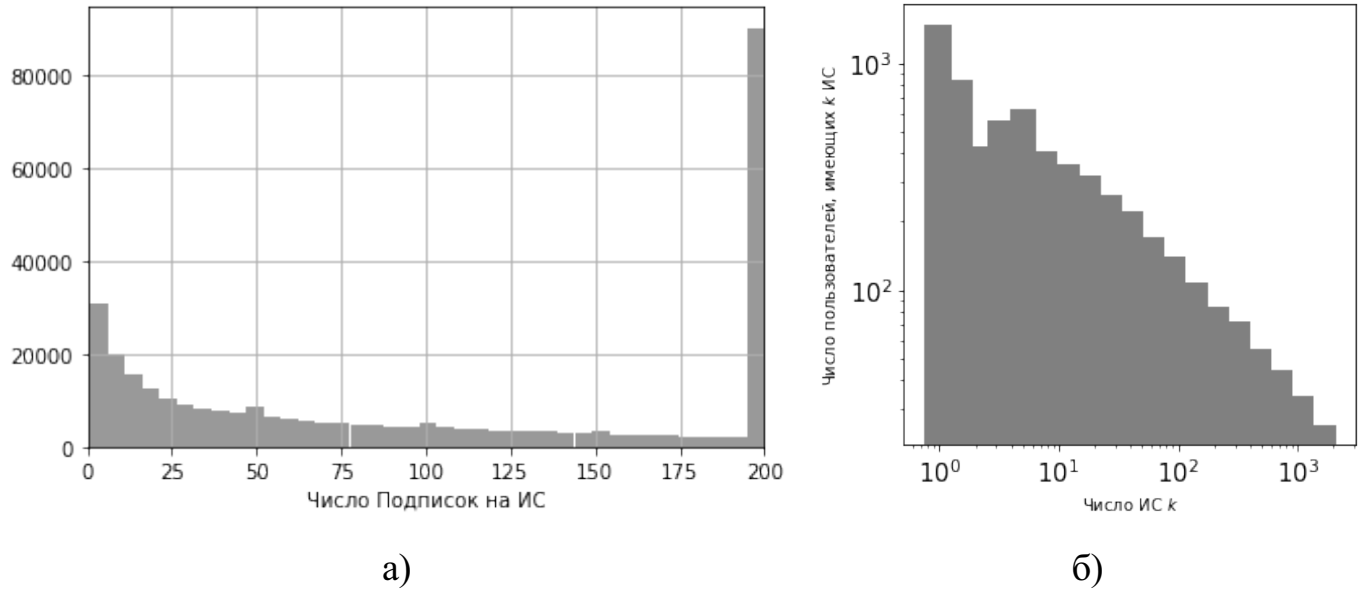


Рисунок 3.3 — Подписки на ИС (а) и подписки на ИС без обрезки сверху в логарифмическом масштабе (б) пользователей Д1

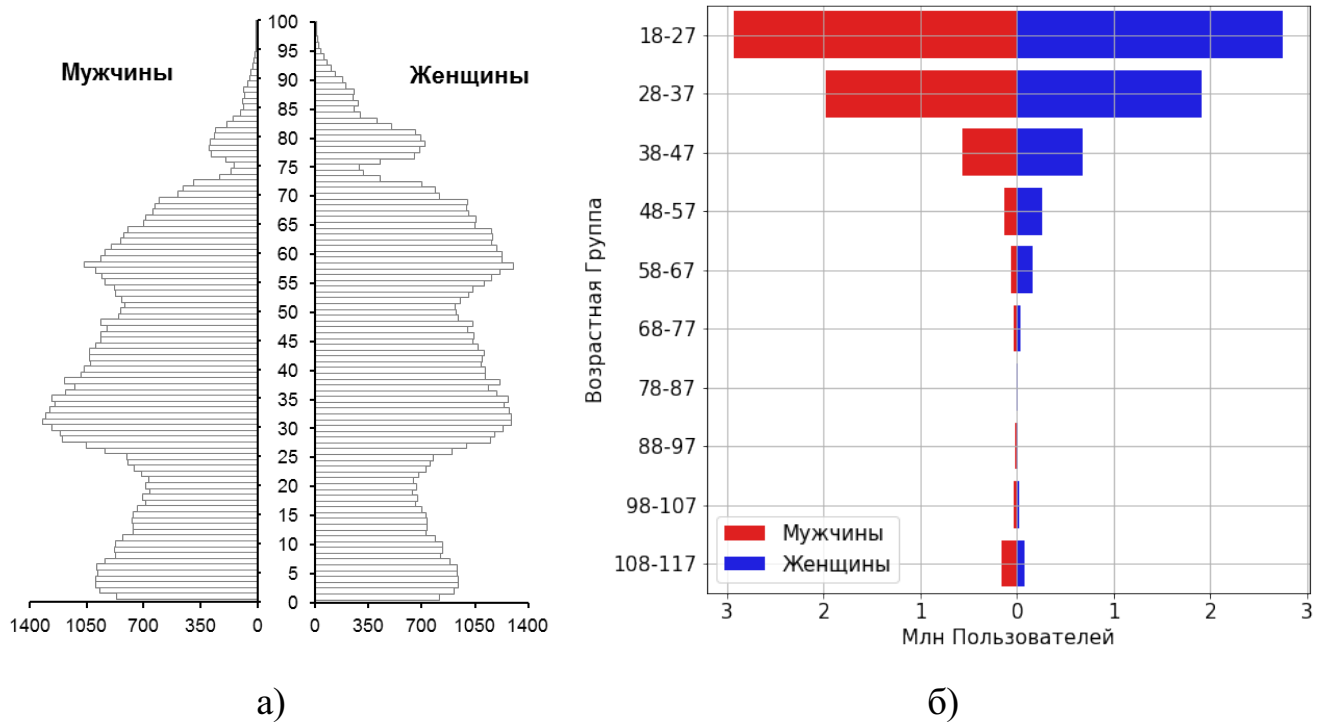


Рисунок 3.4 — Половозрастная пирамида (а) населения Российской Федерации и (б) пользователей Д2.

соответствующее отсутствию информации об объекте). Значение фокального признака в one-hot кодировке представляется единичным вектором

$$\left[0, \dots, \underbrace{1}_i, \dots, 0 \right]^T \in \mathbb{R}^k,$$

где единица находится на i - й позиции, если данный признак принимает i - е значение в своем изначальном представлении. В случае ИС в качестве значений признака выступает факт подписки на данную ИС (подписан/не подписан), а в случае ПП – одна из 10 альтернатив. При этом, если всего возможных уникальных значений ИС k , то тогда возможных значений признака – 2^k . Обозначим число выгруженных в Д1 пользователей как N_{D1} .

ИС и ПП пользователей были закодированы в разреженные csr-матрицы $\mathbf{X}_{IS} \in \mathbb{R}^{N_{D1} \times N_{IS}}$ и $\mathbf{X}_{political} \in \mathbb{R}^{N_{D1} \times N_{political}}$, соответственно, где $N_{IS} = 2109387$ и $N_{political} = 10$. Размерность пространства признаков после кодировки составила $N_f = N_{IS} + N_{political}$ — это количество уникальных ИС среди пользователей Д1 (N_{IS}) плюс число возможных значений поля ПП ($N_{political}$). Пол и возраст пользователей также были закодированы с помощью one-hot кодирования в csr-матрицы $\mathbf{X}_{age} \in \mathbb{R}^{N_{D1} \times 10}$ и $\mathbf{X}_{sex} \in \mathbb{R}^{N_{D1} \times 3}$. Число столбцов матрицы \mathbf{X}_{sex} соответствует числу уникальных значений характеристики "пол":

- Мужской.
- Женский.
- Нет информации.

При кодировке возраста, рассматривались следующие десять промежутков: $(-\infty, 0]$ (этот промежуток соответствует пользователям, которые не указали свой возраст), $(0-18]$, $(18-23]$, $(23-26]$, $(26-36]$, $(36-45]$, $(45-51]$, $(51-56]$, $(56-86]$, $(86, +\infty)$.

Все закодированные данные сохранены как разреженные csr-матрицы.

Кодирующий алгоритм обучался на Д1, а затем использовался для обработки Д2. Таким образом, он “знал” только те значения признаков, которые были представлены в Д1.

3.2.6 Процедура отбора классификаторов

Для визуализации качества работы классификатора использовалась нормированная матрица ошибок (которая для простоты называется просто матрицей ошибок). В качестве метрики качества применялась (средняя) полнота, которая в случае сбалансированной выборки совпадает с верностью.

Первоначальный подход состоял в том, чтобы разделить Д1 на тренировочную и тестовую выборки и при выборе модели классификатора ориентироваться на его полноту на тестовой выборке.

Однако, в случае если проделать эти операции, то, в силу алгоритма проектирования Д1, любой классификатор переобучится на маркерных аккаунтах и покажет практически максимальную полноту (см. рисунок 3.5).

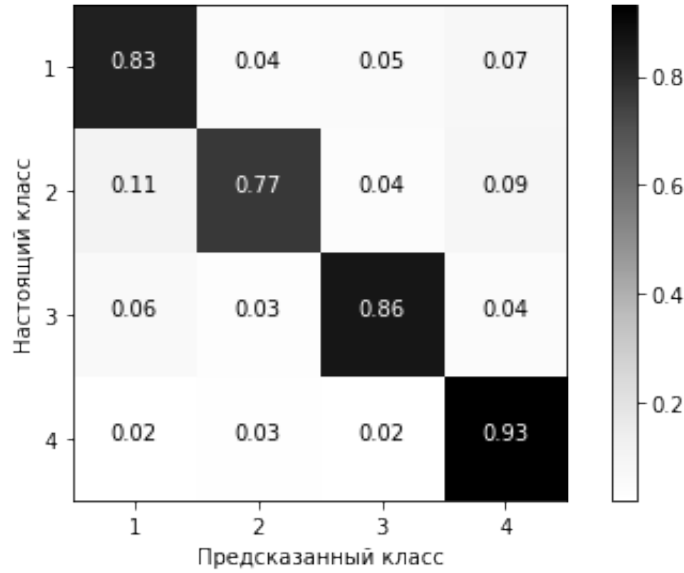


Рисунок 3.5 — Типичный вид матрицы ошибок, когда и обучение, и тестирование происходит только на изначальных пользователях Д1.

В связи с этим, Д1 был дополнен N_{D_1} искусственно сгенерированными аккаунтами, которые являлись полными копиями аккаунтов Д1 за исключением того, что у новых аккаунтов были обнулены компоненты, отвечающие за подписи на маркерные аккаунты. То есть, к матрице объекты-признаки

$$\mathbf{X}_{full} = \begin{bmatrix} \mathbf{X}_{IS} & \mathbf{X}_{political} \end{bmatrix} \in \mathbb{R}^{N_{D_1} \times N_f}$$

и вектору лейблов \mathbf{y} были добавлены матрица объекты-признаки

$$\mathbf{X}_{trunc} = \begin{bmatrix} \mathbf{X}_{IS\ trunc} & \mathbf{X}_{political} \end{bmatrix} \in \mathbb{R}^{N_{D_1} \times N_f}$$

и вектор лейблов

$$\mathbf{y}_{trunc} := \mathbf{y},$$

описывающие признаки и классы синтетических аккаунтов. При этом, в матрице $\mathbf{X}_{IS\ trunc}$ обнулены все компоненты, соответствующие подпискам на маркерные аккаунты.

Пример 3.2.1. Пусть число уникальных пабликов равняется 10-ти, при этом, 1-й и 5-й являются маркерными (будем называть их $M1$ и $M2$ соответственно). Предположим, что матрица \mathbf{X}_{IS} имеет вид:

$$\mathbf{X}_{IS} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

Из матрицы \mathbf{X}_{IS} следует, что 1-й пользователь подписан на $M1$, второй - на $M2$, а третий и на $M1$, и на $M2$. В таком случае, матрица $\mathbf{X}_{IS\ trunc}$ будет иметь следующий вид:

$$\mathbf{X}_{IS\ trunc} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

Используя одно и то же зерно `random state=42` для генерации случайного разбиения, данные

$$\begin{bmatrix} \mathbf{X}_{full} & | & \mathbf{y} \end{bmatrix}$$

и

$$\begin{bmatrix} \mathbf{X}_{trunc} & | & \mathbf{y} \end{bmatrix}$$

были разделены на две обучающие

$$\begin{bmatrix} \mathbf{X}_{full\ train} & | & \mathbf{y}_{full\ train} \end{bmatrix}, \begin{bmatrix} \mathbf{X}_{trunc\ train} & | & \mathbf{y}_{trunc\ train} \end{bmatrix}$$

и две тестовые

$$\begin{bmatrix} \mathbf{X}_{full\ test} & | & \mathbf{y}_{full\ test} \end{bmatrix}, \begin{bmatrix} \mathbf{X}_{trunc\ test} & | & \mathbf{y}_{trunc\ test} \end{bmatrix}$$

выборки (см. рисунок 3.5). При этом, выполнено $\mathbf{y}_{full\ train} = \mathbf{y}_{trunc\ train}$ и $\mathbf{y}_{full\ test} = \mathbf{y}_{trunc\ test}$. Классификаторы обучались на данных

$$\begin{bmatrix} \mathbf{X}_{full\ train} & | & \mathbf{Y}_{full\ train} \\ \mathbf{X}_{trunc\ train} & | & \mathbf{Y}_{trunc\ train} \end{bmatrix},$$

а тестировались на

$$\begin{bmatrix} \mathbf{X}_{trunc\ test} & | & \mathbf{Y}_{trunc\ test} \end{bmatrix}.$$

Таким образом, проверялось, насколько успешно классификаторы действуют в ситуации, когда пользователь не подписан на маркерные аккаунты, что помогало более трезво взглянуть на качество классификации.

```
#CREATE TRAIN AND TEST SAMPLES FOR BOTH FULL AND TRUNCATED PIECES OF DATA
x_full_train, x_full_test, y_full_train, y_full_test = train_test_split(x_full,
                                                                    y,
                                                                    test_size=0.33, random_state=42)
x_trunc_train, x_trunc_test, y_trunc_train, y_trunc_test = train_test_split(x_trunc,
                                                                    y,
                                                                    test_size=0.33, random_state=42)
```

Рисунок 3.6 — Скриншот кода, предназначенного для разделения данных Д1 на обучающие и тестовые выборки.

3.2.7 Отсеивание аномалий

Аномалиями называют объекты, которые своим “нетипичным” поведением “загрязняют” данные, в результате снижая качество классификации [92—94]. В ходе исследований было обнаружено, что пользователи с малым числом подписок на ИС являются таковыми. Причиной тому может быть тот факт, что те, у кого мало подписок на ИС, редко используют ВКонтакте для получения информации, а следовательно, могут исказить искомую зависимость между признаками и метками классов. Был выявлен оптимальный порог аномалий, позволяющий найти баланс между качеством классификации и количеством пользователей, остающимся после удаления аномалий: $th^* = 10$ ИС. Изначальные пользователи, число ИС которых меньше или равно th^* , считались аномалиями и исключались из обучающей и тестовой выборки. Затем выбрасывались синтетические аккаунты, являющиеся производными от исключенных до этого. На рисунке 3.7 и в таблице 3 показано, как аномалии влияют на качество классификации Предсказательной Модели. Из них видно, что чем больше

порог аномалий, тем выше качество классификации. Однако, поскольку распределение числа ИС является степенным (см. рисунок 3.3) слишком высокий порог аномалий может привести к резкому снижению числа объектов в выборке. При использовании порога th^* в выборке оставалось 277756 изначальных и столько же синтетических аккаунтов.

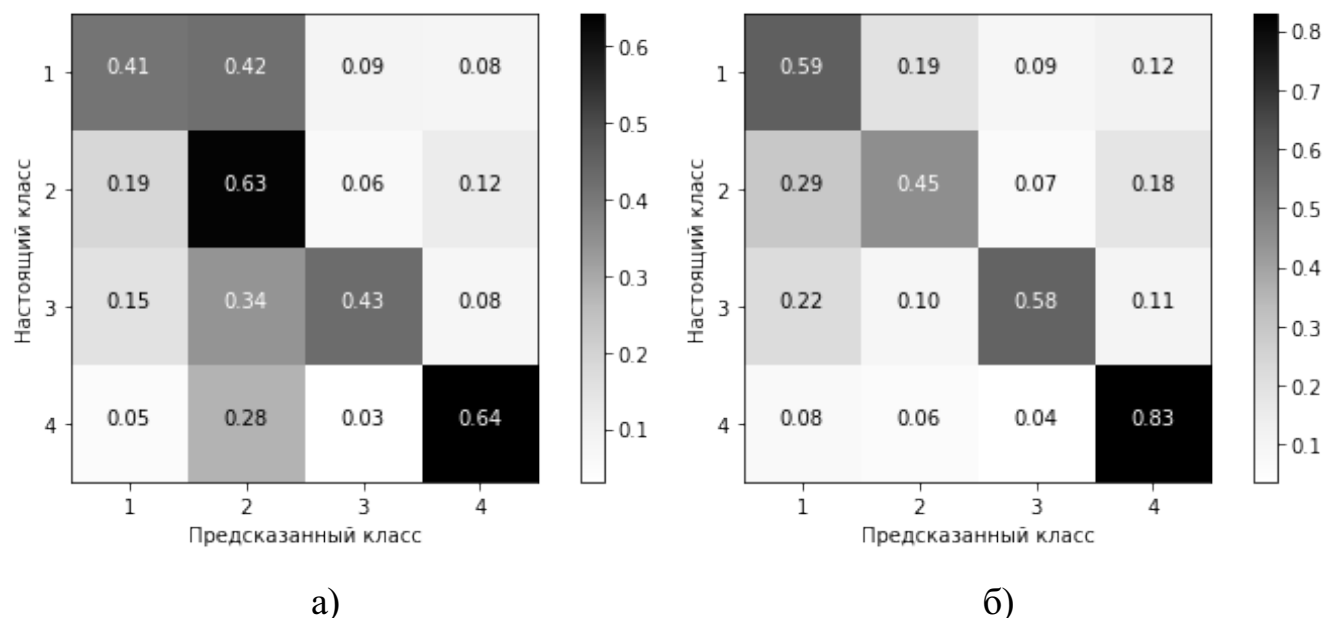


Рисунок 3.7 — Влияние порога аномалий th на качество классификации Предсказательной Модели. (а) $th = -1$ (то есть без отсеивания аномалий), средняя полнота 0.528. (б) $th = 10$, средняя полнота 0.613.

Таблица 3 — Влияние порога аномалий th на качество классификации Предсказательной Модели

Порог аномалий th	Средняя полнота
-1	0.528
0	0.592
5	0.603
10	0.613
15	0.599
20	0.603

3.2.8 Балансировка классов

Несбалансированность классов (представителей класса 4 больше, чем представителей любого из остальных классов) в обучающей выборке может привести к переобучению моделей [92—94]. Вследствие этого использовались стандартные методы балансировки обучающей выборки, в результате которых классы становятся равными по численности. Балансировка проводилась после исключения аномалий. При этом, наилучшее качество классификации было получено при использовании метода под названием оверсэмплинг, в рамках которого миноритарный/ые (меньший по численности) класс/классы насыщаются случайными элементами этого/этих же класса/классов до тех пор, пока не будет достигнут баланс.

3.2.9 Понижение размерности

Для того, чтобы увеличить скорость алгоритмов обучения, а также обойти проблему мультиколлинеарности, от которой страдают линейные классификаторы, пространство признаков подвергалось трансформации, в результате которой его размерность существенно снижалась.

Процедура понижения размерности производилась непосредственно после балансировки классов и перед обучением классификатора.

Применялся метод, основанный на сингулярном разложении (SVD) [94]. Более точно, использовалась его модификация, предназначенная для работы с разреженными матрицами – Truncated SVD (<https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.TruncatedSVD.html>, <https://scikit-learn.org/stable/modules/decomposition.html#lsa>). Ее отличие от метода главных компонент состоит в том, что перед началом работы она не центрирует данные (поскольку центрирование данных приведет к тому, что разреженная матрица превратится в плотную).

Декомпозитор обучался на данных

$$\mathbf{X}_{train} = \begin{bmatrix} \mathbf{X}_{full\ train} \\ \mathbf{X}_{trunc\ train} \end{bmatrix},$$

а затем применялся для трансформации пространства. Для вычисления собственных чисел и собственных векторов матриц $\mathbf{X}_{train}\mathbf{X}_{train}^T$ и $\mathbf{X}_{train}^T\mathbf{X}_{train}$ применялся метод IRAM (который приближенно находит собственные числа и вектора матрицы, конструируя ортонормированный базис подпространства Крылова). В результате, размерность пространства признаков понижалась до 100 компонент.

3.2.10 Классификатор

Классификаторы выбирались среди существующих моделей машинного обучения. Для вызова тестируемых классификаторов использовались библиотеки scikit-learn, TensorFlow, Keras и XGBoost.

В результате процедуры отбора, был выделен классификатор, в основе которого лежит статистическая модель логистической регрессии (логит модель) [92–94]. Преимуществом данного класса классификаторов является скорость обучения, а также возможность интерпретации результатов и, в силу линейности, относительно низкая вероятность переобучения.

Опишем принцип работы классификатора, основанного на логит модели. Пусть число объектов в обучающей выборке равно N , размерность пространства признаков составляет n_f , матрица объекты признаки есть $\mathbf{X} \in \mathbb{R}^{N \times n_f}$, а вектор лейблов равен $\mathbf{y} \in \{-1, 1\}^N$. При обучении логит модели на рассматриваемой выборке решается следующая оптимизационная задача:

$$\min_{\mathbf{w}, w_0} \sum_{i=1}^N \log_2 \left(\exp \left(-y_i \left(\mathbf{x}_i^T \mathbf{w} + w_0 \right) \right) + 1 \right), \quad (3.3)$$

где $\mathbf{w} \in \mathbb{R}^{n_f}$ и $w_0 \in \mathbb{R}$ – коэффициенты регрессии (последний – свободный член), а $\mathbf{x}_i \in \mathbb{R}^{1 \times n_f}$ – i -я строка матрицы \mathbf{X} . Оптимизационная задача решалась с помощью алгоритма CD (Coordinate Descent) https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html.

Как только коэффициенты регрессии найдены, появляется возможность оценить апостериорную вероятность принадлежности i -го объекта каждому из классов:

$$P\{y_i = 1 \mid \mathbf{x}_i\} = \frac{1}{1 + \exp(\mathbf{x}_i^T \mathbf{w} + w_0)} = 1 - P\{y_i = -1 \mid \mathbf{x}_i\}. \quad (3.4)$$

Основная, 4-х классовая задача классификации была сведена к рассмотренной выше бинарной с помощью схемы One-vs-Rest [94], в рамках которой обучалось 4 классификатора, каждый из которых учился отличать объекты соответствующего класса от объектов всех остальных классов в совокупности. При классификации объекта для каждого из четырех классов формировались вероятности (3.4), которые, после соответствующей нормировки, компоновались в вектор (3.1). Таким образом, компоненты вектора (3.1) можно интерпретировать как апостериорные вероятности принадлежности каждому из рассматриваемых классов (поддержки соответствующих фигур). Необходимо отметить, что предложенная трактовка носит достаточно условный характер и должна быть использована с большой осторожностью. С точки зрения теории машинного обучения, приведенные апостериорные вероятности дают нормированную оценку удаленности классифицируемых объектов от границ классов.

Чтобы избежать переобучения, использовалась L2-регуляризация с коэффициентом $C = 0.1$ в рамках которой вместо (3.3) минимизировалась модифицированная функция потерь:

$$C \sum_{i=1}^N \log_2 (\exp(-y_i (\mathbf{x}_i^T \mathbf{w} + w_0)) + 1) + \frac{1}{2} \mathbf{w}^T \mathbf{w}.$$

Данный классификатор, в совокупности с процедурами, которые были описан до этого (выбор признаков, кодировка, создание синтетических аккаунтов, отсеивание аномалий, балансировка классов и понижение размерности) образуют Предсказательную Модель. Результаты, которые показывает Предсказательная Модель на тестовой выборке, приведены на рисунке 3.8.

Исходя из матрицы ошибок, представленной на рисунке 3.8, можно сделать вывод, что Предсказательная Модель достаточно успешно отличает сторонников В.В. Путина от пользователей других классов, в то время как уловить различия между сторонниками П.Н. Грудина, В.В. Жириновского и А.А.

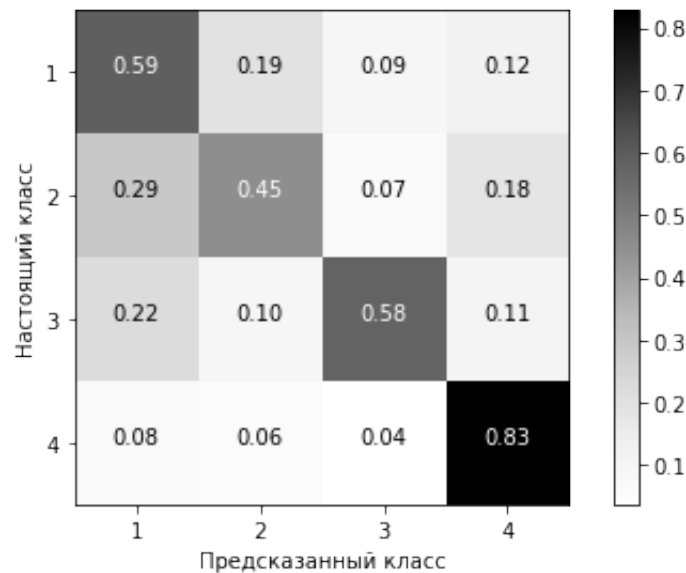


Рисунок 3.8 — Качество классификации Предсказательной Модели на тестовой выборке. Средняя полнота равна 0.613.

Навального – гораздо более трудная задача для нее. Таким образом, если предположить, что что искомая зависимость между независимыми переменными и зависимыми в Д1 и Д2 суть одна и та же и она обнаружена Предсказательной Моделью, то наиболее точным будет прогноз числа пользователей, симпатизирующих В.В. Путину и, следовательно, суммарного числа пользователей, поддерживающих других кандидатов.

3.2.11 Влияние выбора комбинации признаков на качество классификации

Как было отмечено до этого, в качестве независимых переменных, используемых для классификации аккаунтов, использовались ИС, на которые подписан пользователь и выбранные им ПП. При этом, рассматривались другие комбинации независимых переменных, в том числе включающие в себя пол и возраст пользователей. На рисунке 3.9 показано, как зависит качество классификации Предсказательной Модели в зависимости от выбора комбинации независимых переменных. Из рисунка видно, что основной вклад в точность классификации дают ИС, на которые подписан пользователь, при этом пол и возраст пользователей никак не влияет на качество классификации (если рассматривать только три знака после запятой).

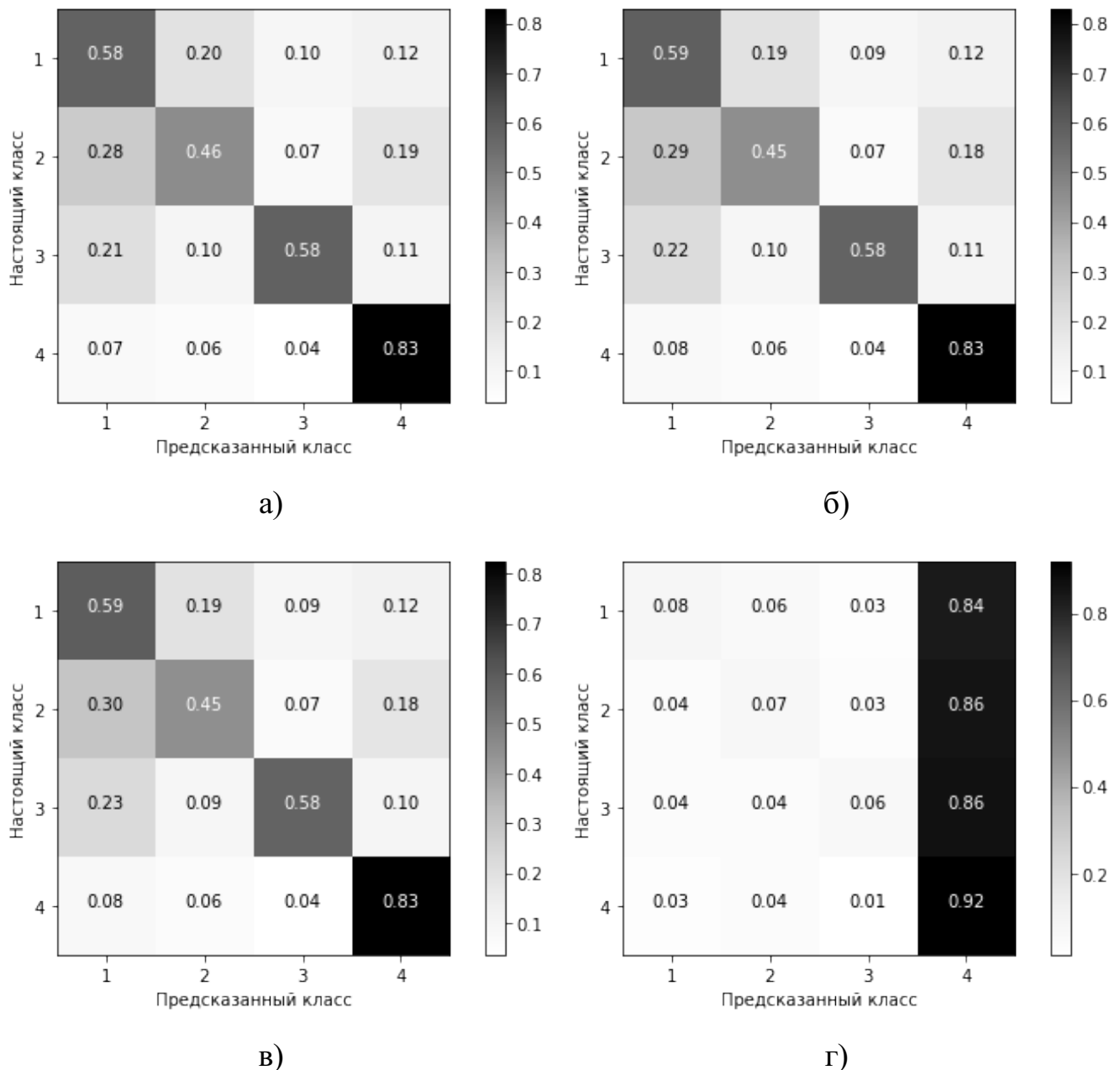


Рисунок 3.9 — Матрица ошибок для случая, когда для построения пространства признаков используются: (а) политические взгляды, ИС, а также пол и возраст; (б) политические взгляды и ИС; (в) только ИС и (г) только политические взгляды. Средняя полнота: 0.613, 0.613, 0.61 и 0.282 соответственно.

3.2.12 Анализ Д2

Обученная на Д1, Предсказательная Модель была применена для классификации аккаунтов Д2. При предобработке данных Д2 использовались те же процедуры, что и в случае Д1 (за исключением балансировки классов, которая

в данном случае не имела смысла). В результате отсеивания аномалий осталось 11820412 аккаунта. Результаты их классификации приведены в таблице 4.

Таблица 4 — Сравнение результатов классификации Д2 с помощью Предсказательной Модели с результатами выборов Президента РФ 2018 года.

	Предсказательная Модель	Итоги выборов
В.В. Путин	0.527	0.767
П.Н. Грудинин	0.256	0.118
В.В. Жириновский	0.123	0.057
А.А. Навальный	0.094	
Другие		0.058

Использование подобных оценок для оценивания итогов выборов не совсем корректно. Например потому, что А.А. Навальный не участвовал в них. Кроме того, необходимо учитывать фактор явки, нерепрезентативность пользователей Вконтакте по отношению к населению страны в целом, а также тот факт, что электоральное поведение не всегда совпадает с симпатиями голосующего. Однако, если интерпретировать полученные показатели как оценку результатов выборов (сравнивая их с официальными результатами), рассматривая тех пользователей Д2, кого Предсказательная Модель определила сторонниками А.А. Навального, как голоса, полученные не рассматриваемыми нами кандидатами в совокупности, то оценка, сделанная Предсказательной Моделью, правильно расставляет первые три места, при этом средняя абсолютная ошибка составляет 0.12.

Функционал ВКонтакте позволяет пользователям в своих профилях указывать (при желании) родной город, а также текущий город проживания. Это позволяет агрегировать результаты классификации по городам и регионам, получая оценку уровня поддержки той или иной политической фигуры в данном конкретном регионе или городе. В таблице 5 приведены пять городов с наибольшими и наименьшими показателями уровня поддержки В.В. Путина. В качестве городов использовались текущие города проживания, имеющие не менее 3000 представителей в Д2.

Таблица 5 — Ранжировка городов пользователей Д2 с учетом уровня поддержки В.В. Путина.

Город	Уровень поддержки В.В. Путина
Каспийск	0.737
Кизляр	0.728
Махачкала	0.717
Хасавюрт	0.71
Владикавказ	0.7
...	...
Королёв	0.466
Москва	0.454
Долгопрудный	0.44
Дубна	0.435
Санкт-Петербург	0.433

3.2.13 Построение одномерного пространства мнений

Числа в (3.1), генерируемые Предсказательной Моделью, а также итоговые классы пользователей можно использовать для калибровки тех параметров моделей формирования мнений, которые отвечают за мнения агентов. В случае (3.1), речь идет о непрерывных пространствах мнений (как, например, в случае модели ДеГроота), а для калибровки дискретных мнений (модель голосующего) можно использовать итоговые классы, к которым отнесены пользователи. Например, можно рассмотреть непрерывное пространство мнений, формализуемое вектором из четырех неотрицательных чисел, в сумме составляющих единицу, которые визуализируют политические предпочтения индивида относительно выбранного политического базиса (П.Н. Грудинин, В.В. Жириновский, А.А. Навальный и В.В. Путин).

Как уже было отмечено, результаты, Предсказательная Модель наиболее точно определяет принадлежность (соответственно, отсутствие принадлежно-

сти) к классу 4 (является сторонником В.В. Путина/ не является). В связи с этим, двумя наиболее достоверными частями вектора (3.1) являются его четвертая компонента \tilde{p}_i^4 , а также $\max \{\tilde{p}_i^1, \tilde{p}_i^2, \tilde{p}_i^3\}$ (поскольку при определении итогового класса классификатор ориентируется на максимальную из четырех вероятностей). В связи с этим, перспективным видится анализ пространства мнений, формализуемого отрезком $[0, 1]$, в котором единица соответствует максимальной поддержке В.В. Путина, ноль — минимальной, а $1/2$ — состоянию максимальной неопределенности.

Для оценки \hat{p}_i мнения p_i агента i (параметра модели) предлагается использовать следующую конструкцию:

$$\hat{p}_i = \frac{\tilde{p}_i^4}{\tilde{p}_i^4 + \max \{\tilde{p}_i^1, \tilde{p}_i^2, \tilde{p}_i^3\}} \in [0, 1]. \quad (3.5)$$

Суммируя все вышесказанное, Предсказательная Модель пользователю i ставит в соответствие число \hat{p}_i от нуля до единицы, выражающее степень поддержки им В.В. Путина. Чтобы подчеркнуть временную привязку, далее будем использовать обозначение $\hat{p}_i(t)$, подразумевая, что данные для построения этой оценки были получены в момент времени t .

3.3 Данные о динамике мнений пользователей и их разведочный анализ

Величины (3.5) были использованы для построения временного ряда оценок мнений Российских пользователей ВКонтакте. Далее для простоты мы будем опускать слово "оценка" кроме тех случаев, когда это и так понятно. Временной ряд, о котором идет речь, состоял из трех снимков мнений $\hat{\mathbf{p}}(t_1)$, $\hat{\mathbf{p}}(t_2)$ и $\hat{\mathbf{p}}(t_3)$. Каждый снимок — это вектор размерности $N_u = 1660927$, состоящий из мнений (3.5) пользователей из выборки U :

$$\hat{\mathbf{p}}(t_k) = \begin{bmatrix} \hat{p}_1(t_k) & \dots & \hat{p}_i(t_k) & \dots & \hat{p}_{N_u}(t_k) \end{bmatrix}^T \in \mathbb{R}^{N_u}.$$

Данные, использованные для построения мнений, были выгружены в Феврале (t_1), Июле (t_2) и Декабре (t_3) 2018 года соответственно. В июле 2018 года была выгружена информация о дружеских связях пользователей. Данные были

выгружены и предоставлены лабораторией Наук о больших данных и проблемах общества, ТГУ, Томск, Российская Федерация.

Все три вектора мнений согласованы друг с другом в том смысле, что i -я строка матрицы

$$\hat{\mathbf{P}} = \begin{bmatrix} \hat{\mathbf{p}}(t_1) & \hat{\mathbf{p}}(t_2) & \hat{\mathbf{p}}(t_3) \end{bmatrix} \in \mathbb{R}^{N_u \times 3} \quad (3.6)$$

описывает динамику мнений пользователя i :

$$\hat{\mathbf{P}}_{i,:} = \begin{bmatrix} \hat{p}_i(t_1) & \hat{p}_i(t_2) & \hat{p}_i(t_3) \end{bmatrix}.$$

Для кодирования идентификаторов пользователей была использована хэш-функция SHA-1.

Процесс формирования выборки пользователей U состоял из следующих шагов. Во-первых, был сформирован список пользователей, аккаунты которых:

- старше 17;
- не удалены, не забанены и не закрыты настройками приватности;
- посещают ВКонтакте хотя бы раз в месяц;
- подписаны не менее, чем на 10 ИС и не более, чем на 200 ИС (принадлежащих пространству признаков Предсказательной Модели).

Во-вторых, для большей достоверности результатов, из описанного выше списка пользователей были исключены те аккаунты, которые входили в список ИС, формирующих пространство признаков Предсказательной Модели. В результате, после дополнительно проведенного случайного сэмплирования, был сформирован список пользователей I . Для каждого пользователя $i \in I$ был получен список его друзей ВКонтакте $\text{Fr}(i)$. Далее рассматривались только те пользователи $U \subset I$, кто имеет хотя бы одного друга из I :

$$U = \{i \in I \mid \text{Fr}(i) \cap I \neq \emptyset\}.$$

Обозначим множество друзей пользователя i из U , которые также принадлежат множеству U как V_i :

$$V_i = U \cap \text{Fr}(i).$$

В силу построения, это множество не является пустым. Далее, для удобства, при $i \in U$ множество V_i будем называть множеством друзей пользователя i . Неориентированный простой невзвешенный граф, чьи вершины соответствуют

пользователям U , а каждая вершина i соединена с вершинами V_i , будем обозначать G .

Матрица \hat{P} и граф G , формирующие в совокупности базу данных ДЗ, были сохранены в виде csv-таблицы и разреженной csr-матрицы соответственно в репозитории [95]. В том же репозитории находится код для воспроизведения результатов, приведенных ниже в этом параграфе и в параграфе 3.6.

На рисунках 3.10 и 3.11 представлены гистограммы, показывающие распределения мнений пользователей U в разные моменты времени, а также распределения изменений их мнений. В таблице 6 показано, как менялось среднее $\frac{1}{N_u} \sum_{i=1}^{N_n} \hat{p}_i(t_k)$, стандартное отклонение мнений пользователей U , а также ассортативность и индекс Джини в рассматриваемый промежуток времени. Положительные значения коэффициента ассортативности свидетельствуют о том, что рассматриваемая сеть является ассортативной: высока вероятность, что пользователи, являющиеся друзьями, имеют схожие мнения. Ассортативность социальных графов – чрезвычайно распространенный феномен [65]. Интерес представляет вопрос, в чем причина этого: тот факт, что пользователи предпочитают общаться только с теми, кто разделяет их мнения [65], или же, возможно, разногласия между ними сглаживаются уже в процессе общения (или же, работают одновременно оба механизма) [96]. Рост уровня ассортативности свидетельствует о снижении уровня разногласий между пользователями, являющимися друзьями. В рамках концепции ассимилятивных моделей социального влияния рост уровня поляризации может быть объяснен, например, наличием двух (или более) влиятельных пользователей с достаточно удаленными друг от друга позициями, которые "притягивают" других пользователей "растягивая" их позиции в пространстве мнений.

Таблица 6 — Динамика изменения среднего, стандартного отклонения σ мнений пользователей U , а также ассортативности и индекса Джини.

	Среднее, σ	Ассортативность	Индекс Джини
t_1	0.48, 0.177	0.1406	0.202
t_2	0.477, 0.179	0.1408	0.205
t_3	0.474, 0.18	0.1413	0.208

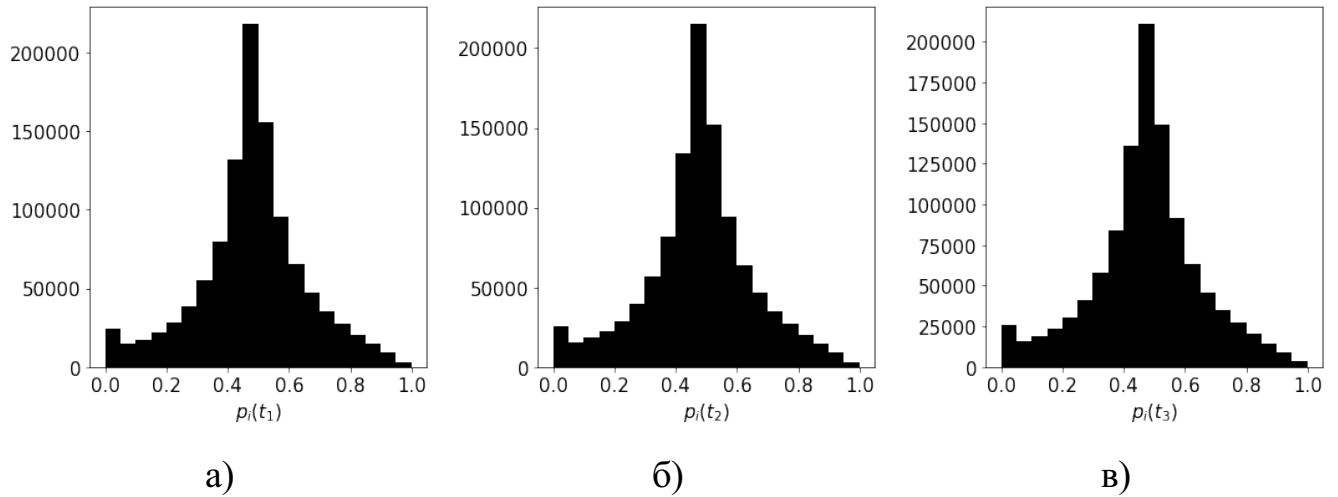


Рисунок 3.10 — Распределения мнений пользователей U в феврале (а), июле (б) и декабре (в) 2018 года.

Распределение степеней вершин графа G_u (см. рисунок 3.12) свидетельствует о том, что степени вершин подчиняются степенному закону:

$$P\{|V_i| = k\} = Ck^{-\alpha}.$$

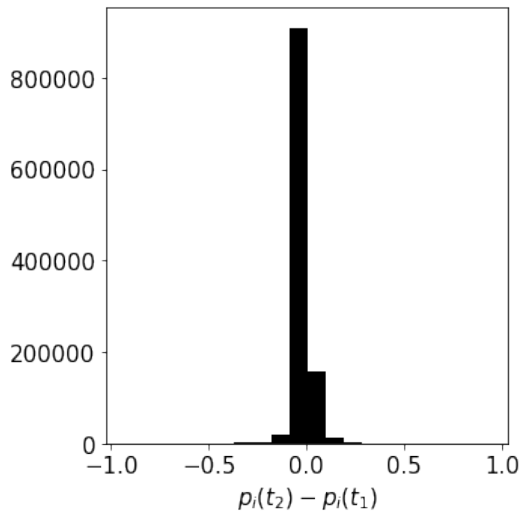
В частности, используя формулу [65, с. 392]

$$\alpha = \frac{N}{\sum_i \ln \frac{k_i}{k_{min}-0.5}}$$

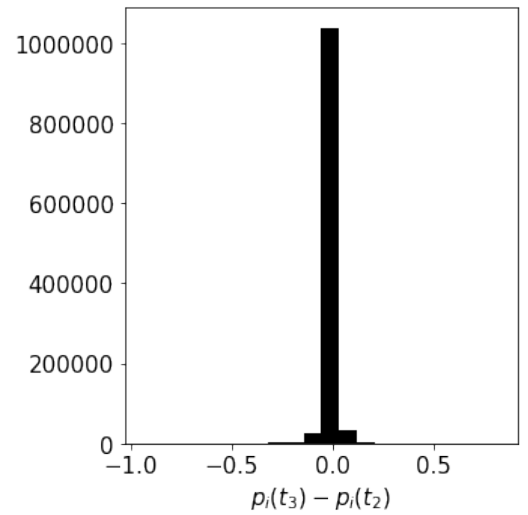
для нахождения показателя степени α (где $k_{min} = 15$ – приблизительная минимальная степень, после которой сохраняется степенной закон, N – количество пользователей, у которых больше или равно k_{min} друзей, и суммирование идет по всем таким пользователям), можно получить

$$\alpha = 2.42.$$

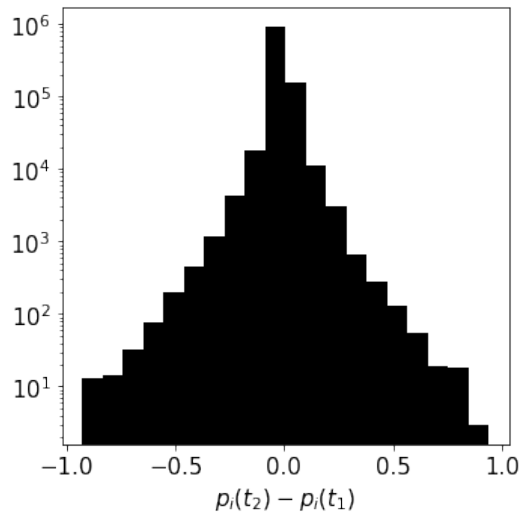
При этом, стандартное отклонение составит 0.003. Стоит отметить, что подобный результат является обыденностью для социальных графов, где α обычно принадлежит промежутку $[2, 3]$ [65].



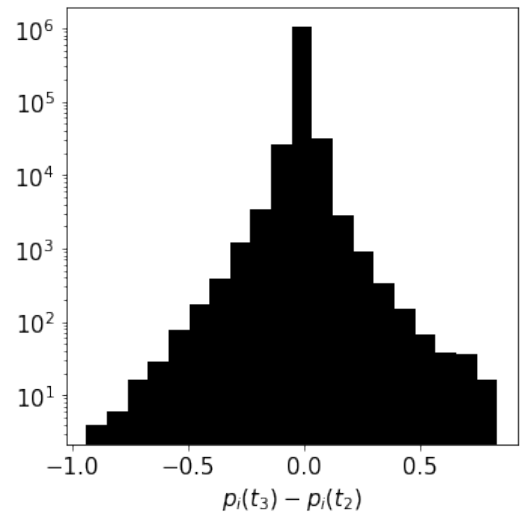
а)



б)



в)



г)

Рисунок 3.11 — Распределения изменений мнений пользователей U в промежутке времени с февраля по июль ((а) - обычная шкала ординат, (в) - логарифмическая) и с июля по декабрь ((б) - обычная шкала ординат, (г) - логарифмическая) 2018 года.

3.4 Аппроксимирующие зависимости

Данные из ДЗ, представленные в параграфе 3.3, были использованы для проверки адекватности моделей Краснощекова и ДеГроота. Однако, в силу малого числа снапшотов – их всего было три – при решении задачи восстановления сети влияния (см. параграф 1.2.4) могла возникнуть проблема переобучения. Поэтому рассматривались *упрощенные* версии этих моделей, для которых про-

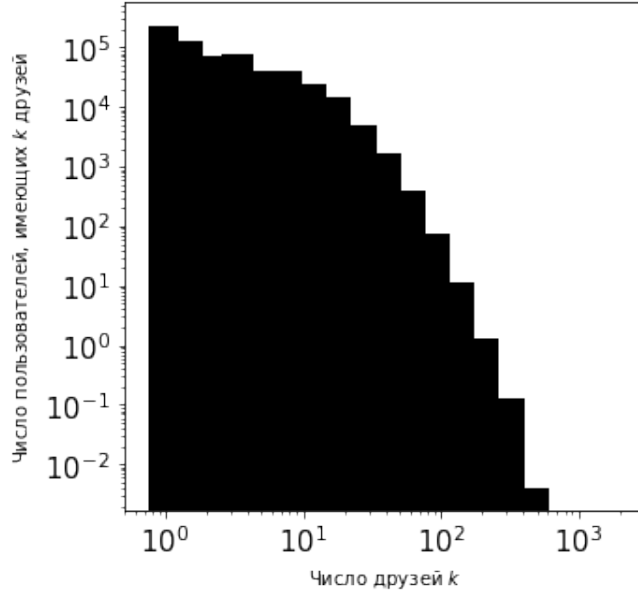


Рисунок 3.12 — Гистограмма распределения степеней вершин графа G_u . Оси абсцисс и ординат имеют логарифмический масштаб. Приблизительная прямолинейная форма средней и хвостовой частей гистограммы (начиная с $k_{min} = 15$) говорит о том, что распределение степеней вершин соответствует степенному закону.

блема переобучения не столь актуальна. Вместе с тем, предиктивная сила таких моделей значительно снижена.

Данные из ДЗ дают возможность для каждого пользователя $i \in U$ посчитать среднее мнение его друзей в момент времени t :

$$\hat{p}_{-i}(t) = \frac{\sum_{j \in V_i} \hat{p}_j(t)}{|V_i|}.$$

На рисунках 3.13 и 3.14 представлены гистограммы, показывающие распределения величин $\hat{p}_{-i}(t_k)$ и $\hat{p}_{-i}(t_{k+1}) - \tilde{p}_{-i}(t_k)$ для $k \in \{1, 2, 3\}$ и $k \in \{1, 2\}$ соответственно.

Для аппроксимации данных использовались следующие зависимости:

$$\hat{p}_i(t_{k+1}) = \frac{1}{|V_i| + 1} (\hat{p}_i(t_k) + |V_i| \hat{p}_{-i}(t_k)), \quad (3.7)$$

$$\hat{p}_i(t_{k+1}) = \frac{1}{|V_i| + 1} (\hat{p}_i(t_k) + |V_i| \hat{p}_{-i}(t_{k+1})), \quad (3.8)$$

$$\hat{p}_i(t_{k+1}) = \mu \hat{p}_i(t_k) + (1 - \mu) \hat{p}_{-i}(t_k), \quad (3.9)$$

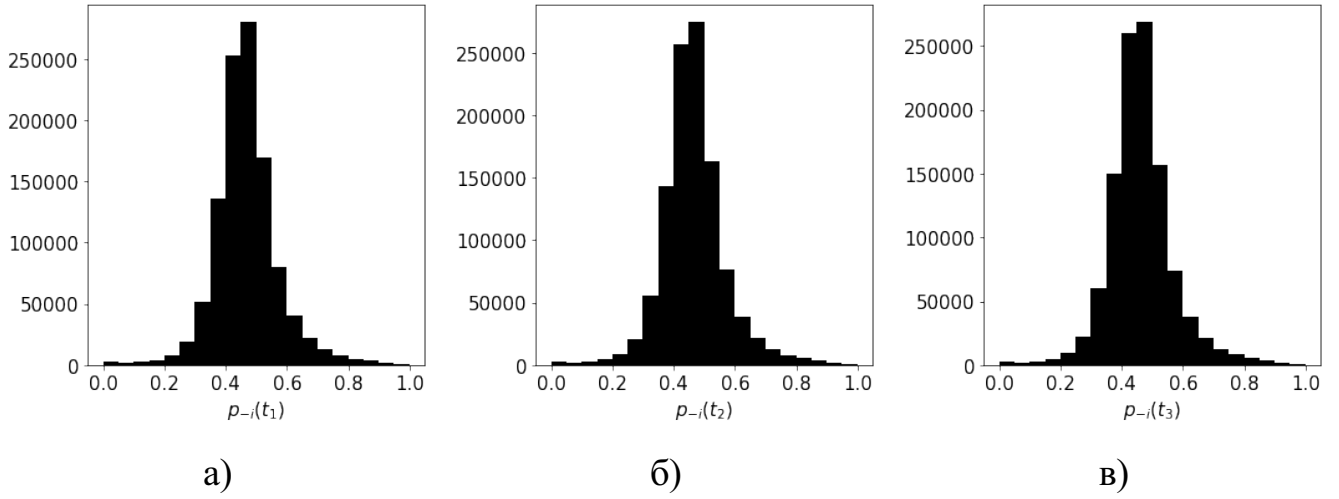


Рисунок 3.13 — Распределения $\hat{p}_{-i}(t_k)$ в феврале (а), июле (б) и декабре (в) 2018 года.

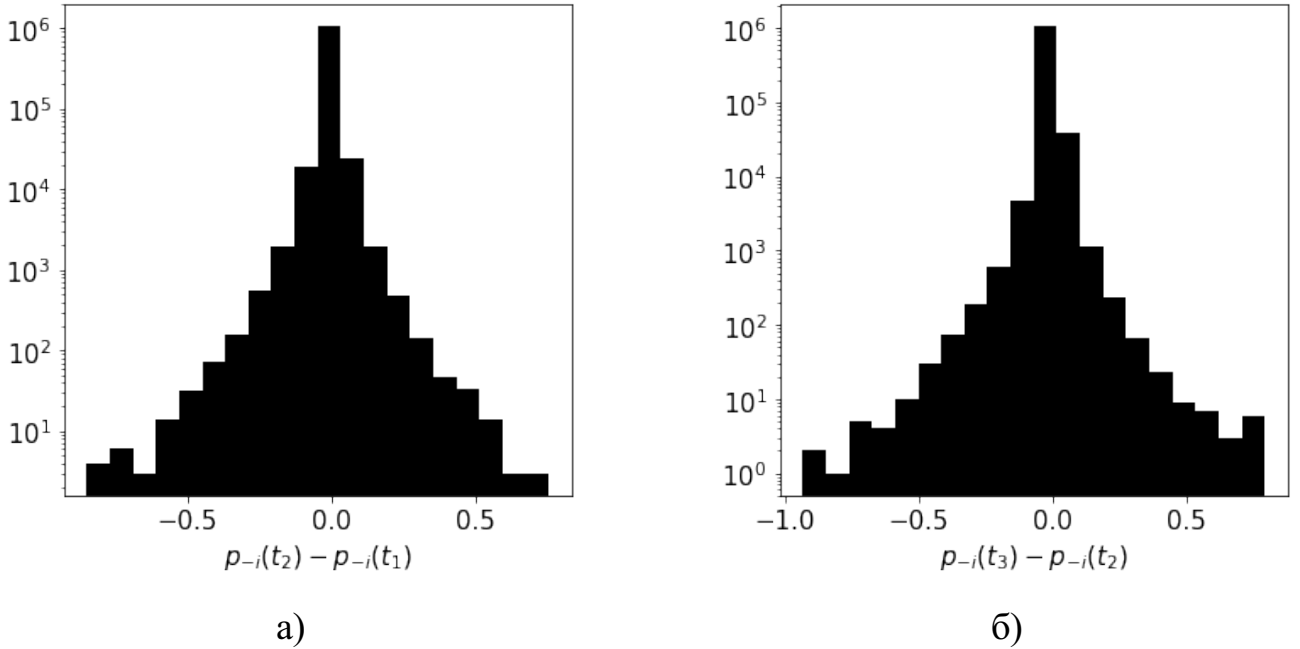


Рисунок 3.14 — Распределения $\hat{p}_{-i}(t_{k+1}) - \hat{p}_{-i}(t_k)$ в промежутке времени с февраля по июль (слева) и с июля по декабрь (справа) 2018 года.

$$\hat{p}_i(t_{k+1}) = \mu \hat{p}_i(t_k) + (1 - \mu) \hat{p}_{-i}(t_{k+1}). \quad (3.10)$$

Выражения (3.7) и (3.9) являются частными случаями итеративных соотношений, описывающих динамику мнений модели ДеГроота, а (3.8) и (3.10) - частными случаями протокола модели Краснощекова. Во всех этих соотношениях влияния агентов $j \in V_i$ на агента i равны друг другу. При этом, в (3.7) и (3.8) влияние агента i на себя (степень сопротивления внешнему влиянию) зависит от числа агентов, имеющих влияние на агента i (от числа его друзей) и

равна влиянию произвольного агента $j \in V_i$ на него. Напротив, в (3.9) и (3.10) влияние агентов на себя являются фиксированной величиной μ , не зависящей от числа друзей.

Кроме того, необходимо отметить, что соотношения (3.7) – (3.10) подразумевают, что сеть влияния между пользователями остается неизменной, что в реальных условиях маловероятно.

Вопрос, на который необходимо ответить изначально выглядел следующим образом: «насколько хорошо соотношения (3.7) – (3.10) соответствует имеющимся данным данным ДЗ? » Этот вопрос был переформулирован: «Насколько хорошо соотношения

$$\hat{p}_i(t_{k+1}) - \hat{p}_i(t_k) = \frac{|V_i|}{|V_i| + 1} (\hat{p}_{-i}(t_k) - \hat{p}_i(t_k)), \quad (3.11)$$

$$\hat{p}_i(t_{k+1}) - \hat{p}_i(t_k) = \frac{|V_i|}{|V_i| + 1} (\hat{p}_{-i}(t_k) - \hat{p}_i(t_{k+1})), \quad (3.12)$$

$$\hat{p}_i(t_{k+1}) - \hat{p}_i(t_k) = (1 - \mu) (\hat{p}_{-i}(t_k) - \hat{p}_i(t_k)), \quad (3.13)$$

$$\hat{p}_i(t_{k+1}) - \hat{p}_i(t_k) = (1 - \mu) (\hat{p}_{-i}(t_{k+1}) - \hat{p}_i(t_k)), \quad (3.14)$$

описывают изменения мнений пользователей U ? »

При этом, важно понимать, что данные из ДЗ являются неполными: не все граждане РФ пользуются социальной сетью ВКонтакте, а те, кто пользуется, могут скрывать свои аккаунтами настройками приватности. Кроме того, при конструировании выборки U проводилось случайное сэмплирование (см. параграф 3.3), что также увеличило степень неполноты.

3.5 Дизайн экспериментов

В экспериментах проводились сравнения левых и правых частей соотношений (3.11) – (3.14), которые рассчитывались на основании имеющихся данных. Сравнения проводились по двум направлениям. Во первых, проверялось, равны ли знаки этих величин. Другими словами, анализировалось,

насколько хорошо соотношения (3.11) – (3.14) предсказывают *направления* изменений мнений пользователей. Во-вторых, левые и правые части (3.11) – (3.14) сравнивались по абсолютной величине. Целью данного сравнения было понять, насколько хорошо модели (3.11) – (3.14) оценивают магнитуды изменений мнений пользователей.

Рассчитывались следующие показатели:

1. Точность определения знака (sign accuracy – SA) изменений мнений пользователей, которая определялась как доля пользователей, направления изменений мнений которых были предсказаны верно;
2. Точность определения знака (correct sign accuracy – CSA) корректных изменений мнений пользователей, которая определялась как доля пользователей, направления изменений мнений которых были предсказаны верно и новые мнения которых не вышли за пределы промежутка, концами которого являются мнение пользователя в предыдущий момент времени и среднее мнение его друзей;
3. Средняя абсолютная ошибка (mean absolute error – MAE) определения магнитуды изменений мнений тех пользователей, чьи направления изменений мнений были предсказаны верно.

Выражения (3.13), (3.14) применялись для аппроксимации данных при различных (но одних и тех же в рамках одного эксперимента) значениях параметра μ . Стоит отметить, что с точки зрения определения SA, разницы между выражениями (3.11) и (3.13), а также (3.12) и (3.14) нет.

При первичном анализе данных было выявлено, что значения SA сильно зависят от $\hat{p}_i(t_{k+1}) - \hat{p}_i(t_k)$. В связи с этим, для наглядного представления результатов полуинтервал $(0, 1]$ был разбит на 10 промежутков вида $(a, b]$ равной длины. Каждый из пользователей с ненулевой магнитудой изменения мнения попадал в один из таких промежутков. В результате, пользователи группировались на основании магнитуд изменений их мнений, после чего для пользователей, попавших в один полуинтервал можно было рассчитывать необходимые показатели. Данное разбиение производилось для каждой из двух итераций независимо.

3.6 Результаты аппроксимации

На рисунке 3.15а представлен типичный вид зависимости SA от магнитуды изменения мнения при аппроксимации данных ДЗ соотношениями (3.11) – (3.14). Соответствующие кардинальные числа множеств пользователей, построенных в соответствии с группировкой пользователей относительно магнитуд изменений их мнений приведены в таблице 7 приложения А. На рисунке 3.15б показано, как CSA зависит от магнитуды изменения мнения.

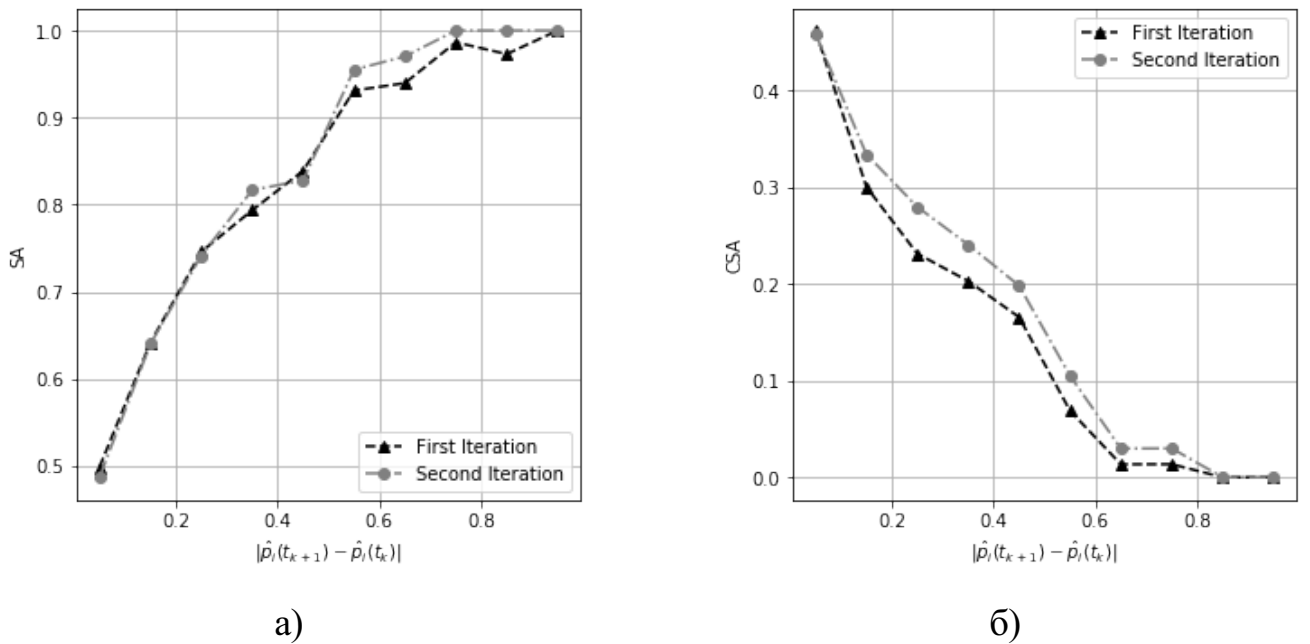
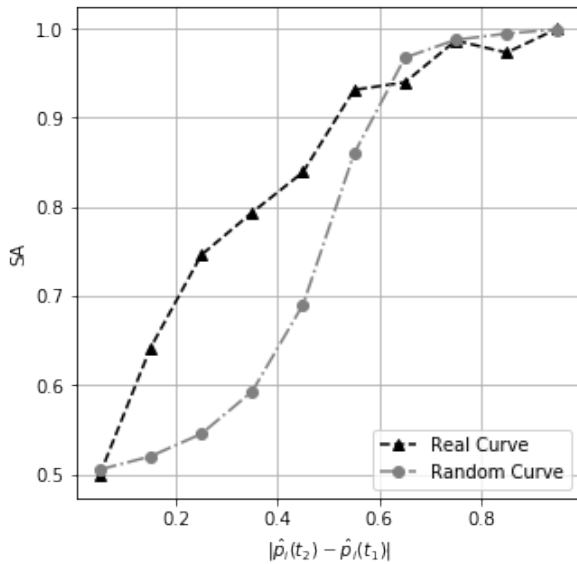


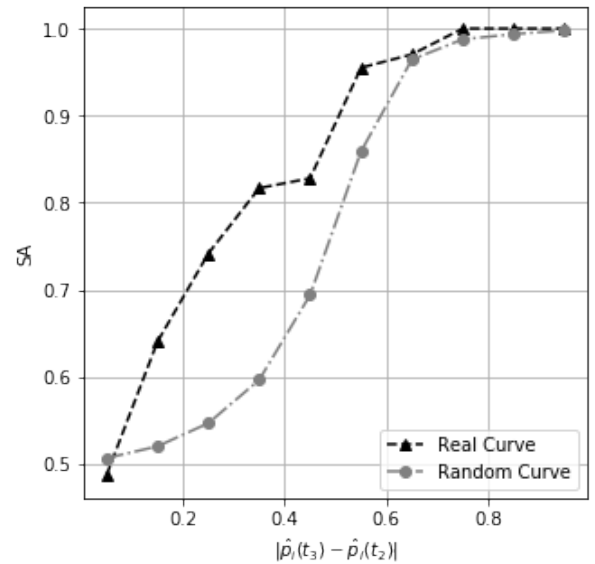
Рисунок 3.15 — SA (слева) и CSA (справа) как функция магнитуды изменения мнения $|\hat{p}_i(t_{k+1}) - \hat{p}_i(t_k)|$.

Для проверки неслучайности результатов, представленных в данном параграфе, проводился следующий эксперимент. Для каждого $k \in \{1, 2\}$ каждому пользователю i был присвоен параметр $\hat{p}_i^*(t_{k+1}) \in [0, 1]$, сгенерированный как случайная величина, равномерно распределенная на отрезке $[0, 1]$. Рассматривая эти параметры как новые мнения пользователей, для каждой из двух итераций были построены соответствующие зависимости. На рисунках 3.16, 3.17 и 3.18 показано, как SA, CSA и MAE зависят от магнитуды изменения мнения в случаях реальных и случайно сгенерированных данных.

На рисунке 3.19 представлена зависимость коэффициента ранговой корреляции Спирмена между величинами $|0.5 - \hat{p}_i(t_k)|$ и $\frac{\hat{p}_i(t_{k+1}) - \hat{p}_i(t_k)}{\hat{p}_{-i}(t_{k+1}) - \hat{p}_i(t_k)}$ (SRCC) от магнитуды изменения мнения.



а)



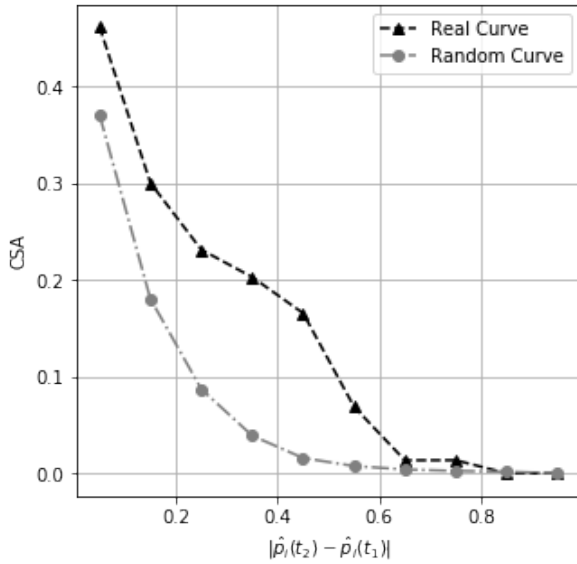
б)

Рисунок 3.16 — Реальная зависимость SA от магнитуды изменения мнения (черные кривые) и в случае случайно сгенерированных изменений мнений пользователей (серые кривые) для первой итерации (слева) и второй итерации (справа).

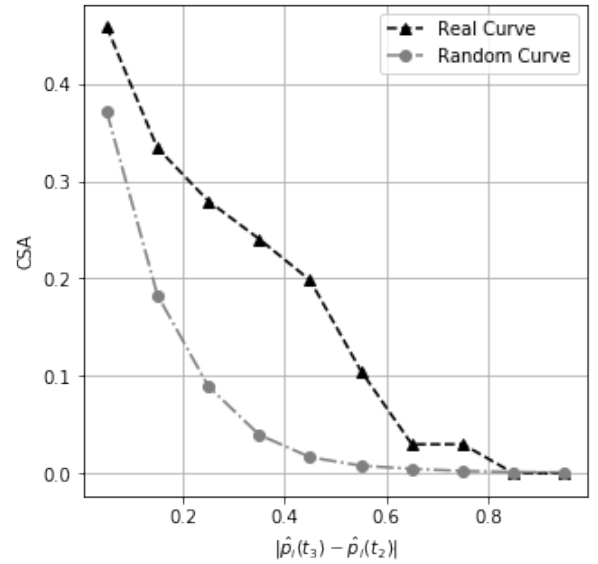
3.7 Анализ результатов

Рисунок 3.15а свидетельствует о том, что SA является монотонно возрастающей функцией магнитуды изменения мнения при аппроксимации данных соотношениями (3.11) – (3.14). Другими словами, чем больше величина изменения мнения, тем лучше соотношения (3.11) – (3.14) выявляют направление изменения этого самого мнения. Более того, общая закономерность такова, что если величина изменения оценки мнения больше 0.4, то тогда SA больше 0.8, что является чрезвычайно высоким показателем. Для малых значений $|\hat{p}_i(t_{k+1}) - \hat{p}_i(t_k)|$ соотношения (3.11) – (3.14) не в состоянии предсказывать направления изменений мнений пользователей. При этом, среди зависимостей (3.11) – (3.14) трудно выделить лидера с точки зрения метрики SA: нет значимой разницы, какую величину использовать: $\hat{p}_{-i}(t_k)$ или $\hat{p}_{-i}(t_{k+1})$.

Однако, из графика, представленного на рисунке 3.15б следует, что значительную долю верно угаданных траекторий пользователей составляют траектории, в рамках которых пользователи не просто меняли свое мнение в сторону мнений их друзей, а даже "перепрыгивали" последние, что, вообще го-



а)

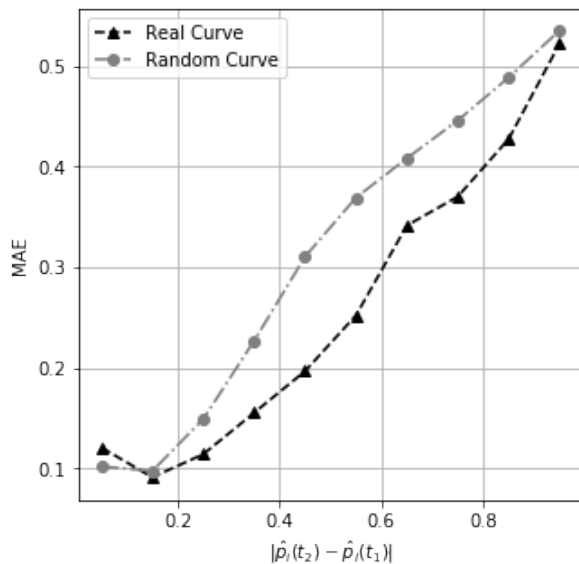


б)

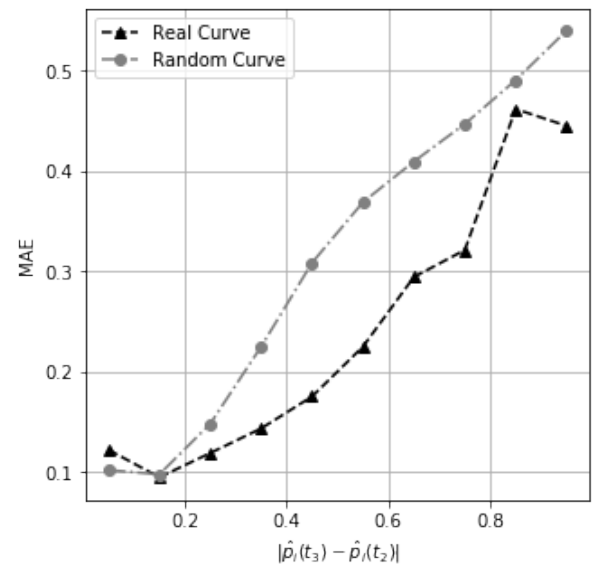
Рисунок 3.17 — Реальная зависимость CSA от магнитуды изменения мнения (черные кривые) и в случае случайно сгенерированных изменений мнений пользователей (серые кривые) для первой итерации (слева) и второй итерации (справа).

вора, противоречит концепции выпуклого механизма формирования взглядов. В связи с этим возникает следующий вопрос. Быть может, высокие значения SA для пользователей, чьи магнитуды изменений мнений достаточно велики, были заранее преопределены? Интуиция за этим следующая. Рассмотрим некоторого пользователя i , для которого выполнено, к примеру, $|\hat{p}_i(t_2) - \hat{p}_i(t_1)| > 0.9$. В силу того, что траектории мнений пользователей не могут выходить за пределы отрезка $[0, 1]$, это означает, что для пользователя i выполнено $\hat{p}_i(t_1) < 0.1$ или $\hat{p}_i(t_1) > 0.9$. Однако в таком случае велика вероятность, что пользователь i поменяет свое мнение таким образом, что изменение его мнения будет направлено в сторону среднего мнения его друзей $\hat{p}_{-i}(t_1)$ будет велика просто потому, что для этого достаточно, чтобы выполнялось $\hat{p}_{-i}(t_1) \in [0.1, 1]$ или $\hat{p}_{-i}(t_1) \in [0, 0.9]$ соответственно.

Чтобы оценить влияние описанных выше факторов на итоговый результат и понять, не являются ли он следствием случайного стечения обстоятельств, зависимость SA от магнитуды изменения мнения, а также другие метрики строились для случая, когда мнения пользователей двигаются случайно по отрезку $[0, 1]$. В рамках такого подхода, пользователь i , имеющий, к примеру, $\hat{p}_i(t_1) < 0.1$, с вероятностью 0.9 сдвинется вправо, то есть, скорее всего, изме-



а)



б)

Рисунок 3.18 — Реальная зависимость MAE от магнитуды изменения мнения (черные кривые) и в случае случайно сгенерированных изменений мнений пользователей (серые кривые) для первой итерации (слева) и второй итерации (справа). Выбор различных значений параметра μ в соотношениях 3.13, 3.14 не изменяет общий характер зависимости.

нит свою позицию в сторону усредненного мнения его друзей (при условии, что последнее также находится справа).

Из графиков, представленных на рисунке 3.16 следует, что полученные результаты для SA не являются исключительно следствием описанной выше закономерности в том смысле, что реальные кривые находятся выше сгенерированных случайным образом при значениях оси абсцисс больше 0.1 и меньше 0.6. Это позволяет сделать вывод, что друзья пользователя действительно оказывают статистически значимое влияние на направление изменения его мнения в случае, если последнее больше 0.1 и меньше 0.6. При этом, однако, рост значения SA при увеличении магнитуды был предопределен заранее, поскольку достигается даже при случайно сгенерированных траекториях.

Что касается небольших по модулю изменений мнений, в любом случае они мало подвержены фактору влияния друзей и, скорее всего, являются артефактом несовершенства Предсказательной Модели, которая, возможно, в состоянии улавливать только значительные изменения мнений пользователей. Кроме того, необходимо учитывать сложную природу человека, поведение которого может быть подвержено различным флуктуациям.

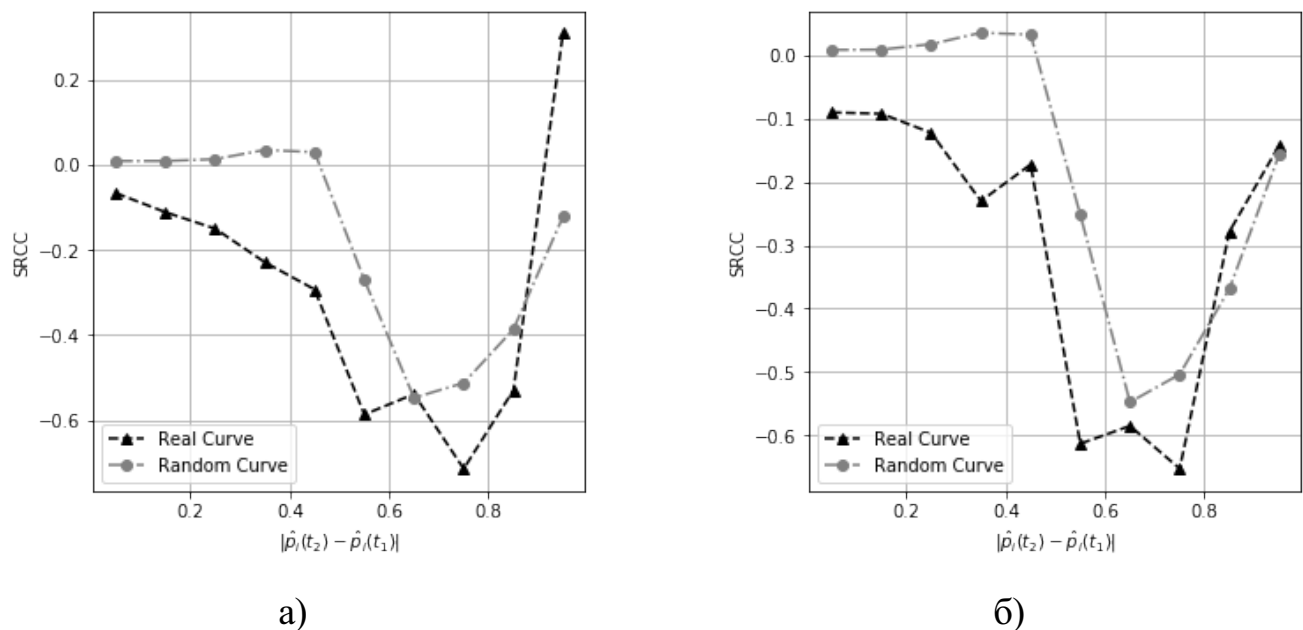


Рисунок 3.19 — Реальная зависимость SRCC от магнитуды изменения мнения (черные кривые) и в случае случайно сгенерированных изменений мнений пользователей (серые кривые) для первой итерации (слева) и второй итерации (справа).

Про изменения мнений пользователей, которые по магнитуде превосходят 0.6 ничего точно сказать нельзя: возможно, они также вызваны влиянием друзей пользователей, однако также и возможно, что причина абсолютно другая.

Рисунок 3.17 позволяет сделать вывод, что, несмотря на чрезвычайно низкие показатели CSA, доля пользователей, направления изменений мнений которых верно определены соотношениями (3.11) – (3.14) и которые не просто двигаются в сторону мнений своих друзей, но также и не "перепрыгивают" их больше, чем та, которая была бы, если бы мнения пользователей двигались бы по отрезку $[0, 1]$ случайным образом (за исключением пользователей, чьи магнитуды изменений мнений больше 0.6).

Рассмотрим зависимость MAE от магнитуды изменения мнения пользователя. Из графика, представленного на рисунке 3.18 следует, что средняя абсолютная ошибка пропорциональна величине изменения мнения. То есть, чем тем сильнее пользователь меняет свое мнение, тем больше ошибка прогнозирования величины этого изменения при использовании (3.11) – (3.14) для построения прогноза. Данный результат свидетельствует о том, что те простейшие соотношения, которые использовались для аппроксимации данных, не в состоянии предсказать величину, на которую изменится мнение пользовате-

ля и для этого, возможно, необходимо использовать более сложные модели, учитывающие гетерогенность поведенческих характеристик пользователей, их изменчивость во времени, а также различные внешние воздействия, оказывающие влияние на рассматриваемую систему. Тем не менее, полученные значения MAE оказались ниже, чем в случае, когда мнения пользователей двигались по отрезку $[0, 1]$ случайным образом.

Основной вывод, который можно сделать из представленных результатов состоит в том, что реальные траектории мнений пользователей больше соответствуют моделям (3.11) – (3.14), чем сгенерированные описанным выше случайным образом. Данный результат можно рассматривать как новое эмпирическое подтверждение феномена конформизма в целом и симметричного выпуклого механизма формирования мнения в частности [9; 16; 58]. Тем не менее, необходимо иметь в виду что рассматриваемые данные являются неполными в силу ряда причин неполными (источники влияния вне Вконтакте, сэмплирование, настройки приватности) и зашумленными (несовершенство Предсказательной Модели, боты, виртуалы). Кроме того, нужно учитывать, что зависимости, которые использовались для аппроксимации данных основывались на предположении о гомогенности и неизменности поведенческих характеристик пользователей, что существенно влияет на их предиктивную силу.

В частности, необходимо иметь в виду, что у пользователей могут быть разные параметры μ_i , отвечающие за их сопротивление к социальному влиянию. В связи с этим возникает проблема их идентификации. Обозначим $1 - \hat{\mu}_i = \frac{\hat{p}_i(t_{k+1}) - \hat{p}_i(t_k)}{\hat{p}_{-i}(t_{k+1}) - \hat{p}_i(t_k)}$. Из графиков, представленных на рисунке 3.19 следует, что

$$\hat{\mu}_i \propto |0.5 - \hat{p}_i(t_k)|$$

Говоря простым языком, чем более уверенная позиция у пользователя, тем труднее заставить его поменять свое мнение. Данный факт достаточно хорошо согласуется с имеющимися результатами в области поведенческой психологии [3].

Идентификацию параметра μ_i пользователя i можно провести в рамках решения следующей задачи выпуклой оптимизации.

$$\begin{cases} \min_{\mu_i} \sum_{k=1}^{k_{max}-1} (\hat{p}_i(t_{k+1}) - \mu_i \hat{p}_i(t_k) - (1 - \mu_i) \hat{p}_{-i}(t_{k+1}))^2, \\ \mu_i \in [0, 1], \end{cases} \quad (3.15)$$

где k_{max} – число имеющихся в распоряжении снимков.

С учетом полученных результатов, выражения (3.11) – (3.14), а точнее соответствующие им модели (ДеГроота и Краснощекова) можно *уточнить*. В силу того, что указанные соотношения описывают динамику мнений пользователей с одинаковым успехом, далее будет идти речь только об одном из них – а именно, об уравнении (3.13). При этом, аналогичные рассуждения можно провести для (3.14). Поскольку далее будет идти речь уже о моделях, мы не будем использовать знак шапки, а вместо слова "пользователь" будем использовать термин "агент":

$$p_i(t_{k+1}) - p_i(t_k) = (1 - \mu) (p_{-i}(t_k) - p_i(t_k)), \quad (3.16)$$

где $p_{-i}(t_k)$ – усредненное мнение агентов, имеющих прямое влияние на агента i . Во-первых, чтобы учесть гетерогенность агентов, необходимо раздать им разные μ_i , заменив (3.16) на

$$p_i(t_{k+1}) - p_i(t_k) = (1 - \mu_i) (p_{-i}(t_k) - p_i(t_k)) \quad (3.17)$$

Соотношение (3.17) предлагается заменить на

$$p_i(t_{k+1}) - p_i(t_k) = (1 - \mu(|0.5 - p_i(t_k)|)) (p_{-i}(t_{k+1}) - p_i(t_k)) + \Theta, \quad (3.18)$$

где $\mu(\cdot)$ – возрастающая функция (явный вид зависимости на данный момент неизвестен и требует дальнейших исследований), а Θ – слагаемое, отвечающее за шум.

Вместе с тем, слагаемое $p_{-i}(t_{k+1})$ можно усложнить, учтя тот факт, что влияния агентов друг на друга неоднородны. Во-первых, это можно сделать в рамках решения модифицированной обратной задачи. Для $i \neq j$ обозначим $l_{ij} = (1 - \mu_i)\lambda_{ij}$ и введем конструкцию $\hat{p}_{-i}^*(t_{k+1}) = \sum_{j \in V_i} l_{ij} \tilde{p}_j(t_{k+1})$. Обратную задачу поставим следующим образом:

$$\begin{cases} \min_{\mu_i, l_{ij}} \sum_{k=1}^{k_{max}-1} (\hat{p}_i(t_{k+1}) - \mu_i \hat{p}_i(t_k) - \hat{p}_{-i}^*(t_{k+1}))^2, \\ \mu_i + \sum_{j \in V_i} l_{ij} = 1, \\ \mu_i \geq 0, l_{ij} \geq 0. \end{cases} \quad (3.19)$$

где k_{max} – число имеющихся в распоряжении снапшотов, а $l_{ij} = (1 - \mu_i)\lambda_{ij}$. Отметим, что размерность задачи выпуклой оптимизации (3.19) существенно снижена за счет рассмотрения только тех влияний, которые обусловлены наличием (видимых наблюдателю) дружеских связей (при $j \notin V_i$ полагаем $l_{ij} = 0$).

В качестве ad-hoc поправки для функции потерь в (3.19) можно рассматривать слагаемые $\varepsilon_i(t_{k+1})$, отвечающие за невидимое для наблюдателя влияние (вообще говоря, меняющееся со временем):

$$\left\{ \begin{array}{l} \min_{\mu_i, l_{ij}, \varepsilon_i(t_{k+1})} \sum_{k=1}^{k_{max}-1} (\hat{p}_i(t_{k+1}) - \mu_i \hat{p}_i(t_k) - \hat{p}_{-i}^*(t_{k+1}) - \varepsilon_i(t_{k+1}))^2, \\ \mu_i + \sum_{j \in V_i} l_{ij} + \varepsilon_i(t_{k+1}) \leq 1, \quad k \in [k_{max} - 1], \\ \mu_i \geq 0, l_{ij} \geq 0, \varepsilon_i(t_k) \geq 0, \quad k \in [k_{max} - 1]. \end{array} \right. \quad (3.20)$$

Однако, такой подход неизбежно приведет к переобучению. В целях регуляризации от $\varepsilon_i(t_{k+1})$ можно потребовать независимость от времени (снизив тем самым размерность задачи (3.20) на $k_{max} - 2$):

$$\left\{ \begin{array}{l} \min_{\mu_i, l_{ij}, \varepsilon_i} \sum_{k=1}^{k_{max}-1} (\hat{p}_i(t_{k+1}) - \mu_i \hat{p}_i(t_k) - \hat{p}_{-i}^*(t_{k+1}) - \varepsilon_i(t_{k+1}))^2, \\ \mu_i + \sum_{j \in V_i} l_{ij} + \varepsilon_i \leq 1, \\ \mu_i \geq 0, l_{ij} \geq 0, \varepsilon_i \geq 0. \end{array} \right.$$

или ограничить сверху величину, на которую оно может меняться

$$\left\{ \begin{array}{l} \min_{\mu_i, l_{ij}, \varepsilon_i(t_{k+1})} \sum_{k=1}^{k_{max}-1} (\hat{p}_i(t_{k+1}) - \mu_i \hat{p}_i(t_k) - \hat{p}_{-i}^*(t_{k+1}) - \varepsilon_i(t_{k+1}))^2, \\ \mu_i + \sum_{j \in V_i} l_{ij} + \varepsilon_i(t_{k+1}) \leq 1, \quad k \in [k_{max} - 1], \\ \mu_i \geq 0, l_{ij} \geq 0, \varepsilon_i(t_{k+1}) \geq 0, \\ |\varepsilon_i(t_{k+1}) - \varepsilon_i(t_k)| < \delta, \quad k \in \overline{2, k_{max} - 1}. \end{array} \right.$$

Кроме того, можно ограничить вклад, который слагаемые $\varepsilon_i(t_{k+1})$ дают в функцию потерь:

$$\left\{ \begin{array}{l} \min_{\mu_i, l_{ij}, \varepsilon_i(t_{k+1})} \sum_{k=1}^{k_{max}-1} (\hat{p}_i(t_{k+1}) - \mu_i \hat{p}_i(t_k) - \hat{p}_{-i}^*(t_{k+1}) - \varepsilon_i(t_{k+1}))^2, \\ \mu_i + \sum_{j \in V_i} l_{ij} + \varepsilon_i(t_{k+1}) \leq 1, \quad k \in [k_{max} - 1], \\ \frac{\varepsilon_i(t_k)}{\mu_i + \sum_{j \in V_i} l_{ij} + \varepsilon_i(t_{k+1})} < \delta, \quad k \in [k_{max} - 1], \\ \mu_i \geq 0, l_{ij} \geq 0, \varepsilon_i(t_k) \geq 0, \quad k \in [k_{max} - 1]. \end{array} \right.$$

Стоит отметить, что представленные выше оптимизационные задачи опираются на предположение о неизменности поведенческих характеристик пользователей, что видится маловероятным. Альтернативный подход состоит в оценке неизвестных характеристик пользователей в рамках одной итерации. Например, за счет анализа изменения множества ИС, на которые подписан пользователь и сопоставления с ИС его друзей. Данный вопрос тесно связан с динамикой онлайн сообществ, изучению которой посвящены публикации [97; 98].

Кроме того, можно учитывать активность пользователя по отношению к его друзьям (выражаемую его действиями, например, лайками, комментариями или репостами), которую также можно использовать для оценки матрицы относительных влияний [16]. Еще один метод подразумевает анализ структуры графа дружбы, соединяющего пользователей. К примеру, можно рассматривать связи между только теми пользователями, у кого есть по крайней мере один общий друг.

Дальнейшие исследования в данной области могут быть связаны с:

1. Улучшением качества классификации Предсказательной Модели;
2. Коллекционированием более обширной базы данных, состоящей из большего числа снимков (чем три) и основанной на большей выборке аккаунтов пользователей;
3. Определением оптимального промежутка времени между снятиями снимков;
4. Определением других факторов (помимо $|0.5 - p_i(t_k)|$), влияющих на значение μ_i ;
5. Идентификацией явного вида зависимости μ_i от этих факторов (в том числе от $|0.5 - p_i(t_k)|$).

6. Восстановлением сети влияний с учетом гетерогенности поведенческих характеристик пользователей;
7. Более подробным изучением влияния неполноты и зашумленности данных на итоговый результат при помощи имитационных экспериментов;
8. Анализом других OCC – Twitter, Facebook, Instagram и т.д.

Заключение

Основные результаты работы заключаются в следующем.

1. Построено обобщение модели Краснощекова на случай разреженной матрицы относительных влияний, исследована связь между моделью Краснощекова и моделями ДеГроота и Фридкина-Джонсона, изучен вопрос сходимости мнений в модели Краснощекова, с помощью протокола спорщиков разрешен временной парадокс модели Краснощекова;
2. Разработаны модель, алгоритм и соответствующее программное обеспечение для оценивания политических взглядов российских пользователей ВКонтакте;
3. Сконструирован временной ряд, описывающий динамику мнений выборки российских пользователей ОСС ВКонтакте с февраля по декабрь 2018 года, а также граф социальных связей российских пользователей ВКонтакте;
4. Полученные данные исследованы с точки зрения упрощенных версий модели Краснощекова и модели ДеГроота (основанных на предположении гомогенности относительной сети влияния) на предмет адекватности последних, обнаружены паттерны социального влияния, получено экспериментальное подтверждение данных моделей с некоторыми дополнительными оговорками;
5. Предложены поправки для моделей Краснощекова и ДеГроота, а также направления для дальнейших исследований в данной области.

Автор выражает глубокую благодарность научному руководителю Белолипецкому А. А., а также Чхартишвили А. Г., Губанову Д. А., Федянину Д. Н. и Бабиной Т. С. за поддержку, помощь и обсуждение результатов.

Список сокращений и условных обозначений

$ \cdot $	кардинальное число множества
\mathbb{N}	множество натуральных чисел
$[n]$	множество всех натуральных чисел от единицы до $n \in \mathbb{N}$
\mathbb{R}	множество вещественных чисел
$\mathbb{R}^{n \times m}$	множество вещественных матриц размера n на m
\mathbf{I}_n	единичная матрица размера n на n
$\mathbb{1}_n$	вектор, состоящий из единиц длины n
$\text{diag}(x_1, \dots, x_n)$	диагональная матрица размера n на n с числами x_1, \dots, x_n на главной диагонали
$\ \cdot\ _p$	l_p норма ($p \geq 0$)

Словарь терминов

Неотрицательная матрица – это матрица, все компоненты которой неотрицательны.

Неотрицательная матрица A всегда имеет неотрицательное собственное число, превосходящее по модулю все ее другие собственные числа. Это число называют **числом Фробениуса**.

Стохастический вектор – это неотрицательный вектор, сумма компонент которого равна единице.

Строчно-стохастическая матрица – это неотрицательная квадратная матрица, все строчные суммы которой равны единице.

Строчно-стохастическая матрица W называется **правильной**, если собственное число $\lambda = 1$ является единственным элементом спектра этой матрицы, принадлежащим единичной окружности в комплексной плоскости.

Строчно-стохастическая матрица W называется **регулярной**, если собственное число $\lambda = 1$ является единственным элементом спектра этой матрицы, принадлежащим единичной окружности в комплексной плоскости и, кроме того, является простым корнем характеристического уравнения.

Строчно-субстохастическая матрица – это неотрицательная квадратная матрица, все строчные суммы которой не превосходят единицу.

Неразложимая матрица – это неотрицательная квадратная матрица, которую с помощью одновременной перестановки строк и столбцов нельзя привести к виду

$$\begin{bmatrix} A & B \\ 0 & C \end{bmatrix},$$

где подматрицы A и C являются квадратными.

Нормальной формой неотрицательной квадратной матрицы A называется матрица A' , которую можно получить из матрицы A с помощью одновременной перестановки строк и столбцов и которая имеет вид

$$\mathbf{A}' = \begin{bmatrix} \mathbf{A}_{1,1} & \dots & \mathbf{A}_{1,l} & \mathbf{A}_{1,l} & \dots & \mathbf{A}_{1,k} \\ \mathbf{0} & \dots & \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_{l,l} & \mathbf{A}_{l,l+1} & \dots & \mathbf{A}_{l,k} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{l+1,l+1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & k, k \end{bmatrix}, \quad (3.21)$$

где подматрицы $\mathbf{A}_{1,1}, \dots, \mathbf{A}_{l,l}, \mathbf{A}_{l+1,l+1}, \dots, \mathbf{A}_{k,k}$ расположены на главной диагонали, являются квадратными и неразложимыми.

Закрытой подматрицей нормальной формы (3.21) квадратной матрицы \mathbf{A} называется любая из подматриц $\mathbf{A}_{l+1,l+1}, \dots, \mathbf{A}_{k,k}$.

Путем длины k в простом графе $G = (V, E)$ из вершины $i_1 \in V$ в вершину $i_{k+1} \in V$ называется последовательность вершин i_1, i_2, \dots, i_{k+1} , где для каждого $j \in \{2, \dots, k\}$ вершина $i_j \in V$ и для каждого $j \in \{1, \dots, k\}$ дуга $(j, j+1) \in E$.

Вершина $i \in V$ простого графа $G = (V, E)$ называется **достижимой** из вершины $j \in V$, если существует путь из вершины j в вершину i .

Простой ориентированный граф $G = (V, E)$ называется **сильно связным**, если для любых $i \in V$ и $j \in V$ вершина i достижима из вершины j .

Компонентой сильной связности графа простого ориентированного графа $G = (V, E)$ называется максимальное по включению подмножество V_0 множества его вершин, такое, что для любых $i \in V_0$ и $j \in V_0$ вершина i достижима из вершины j .

Закрытая компонента сильной связности графа $G = (V, E)$ – это компонента его сильной связности недостижимая из любой вершины графа G , не принадлежащей этой компоненте.

Список литературы

1. *Sherif, M.* The psychology of social norms. / M. Sherif. — 1936.
2. *Asch, S. E.* Effects of group pressure upon the modification and distortion of judgments / S. E. Asch, H. Guetzkow // Documents of gestalt psychology. — 1951. — P. 222—236.
3. *Cialdini, R. B.* Social influence: Compliance and conformity / R. B. Cialdini, N. J. Goldstein // Annu. Rev. Psychol. — 2004. — Vol. 55. — P. 591—621.
4. Models of social influence: Towards the next frontiers / A. Flache [et al.] // Journal of Artificial Societies and Social Simulation. — 2017. — Vol. 20, no. 4.
5. *Proskurnikov, A. V.* A tutorial on modeling and analysis of dynamic social networks. Part I / A. V. Proskurnikov, R. Tempo // Annual Reviews in Control. — 2017. — Vol. 43. — P. 65—79.
6. *Proskurnikov, A. V.* A tutorial on modeling and analysis of dynamic social networks. Part II / A. V. Proskurnikov, R. Tempo // Annual Reviews in Control. — 2018. — Vol. 45. — P. 166—190.
7. *Castellano, C.* Statistical physics of social dynamics / C. Castellano, S. Fortunato, V. Loreto // Reviews of modern physics. — 2009. — Vol. 81, no. 2. — P. 591.
8. *Bullo, F.* Perspectives on Network Systems and Mathematical Sociology / F. Bullo, N. E. Friedkin // Uncertainty in Complex Networked Systems. — Springer, 2018. — P. 399—417.
9. *Friedkin, N. E.* A theory of the evolution of social power: Natural trajectories of interpersonal influence systems along issue sequences / N. E. Friedkin, P. Jia, F. Bullo // Sociological Science. — 2016. — Vol. 3. — P. 444—472.
10. *Ravazzi, C.* Learning influence structure in sparse social networks / C. Ravazzi, R. Tempo, F. Dabbene // IEEE Transactions on Control of Network Systems. — 2017. — Vol. 5, no. 4. — P. 1976—1986.
11. *Губанов, Д. А.* Социальные сети: модели информационного влияния, управления и противоборства / Д. А. Губанов, Д. А. Новиков, А. Г. Чхартишвили. — 2010.

12. From classical to modern opinion dynamics / H. Noorazar [et al.] // arXiv preprint arXiv:1909.12089. — 2019.
13. *Weng, L.* Virality prediction and community structure in social networks / L. Weng, F. Menczer, Y.-Y. Ahn // Scientific reports. — 2013. — Vol. 3. — P. 2522.
14. Competition among memes in a world with limited attention / L. Weng [et al.] // Scientific reports. — 2012. — Vol. 2. — P. 335.
15. The science of fake news / D. M. Lazer [et al.] // Science. — 2018. — Vol. 359, no. 6380. — P. 1094—1096.
16. A 61-million-person experiment in social influence and political mobilization / R. M. Bond [et al.] // Nature. — 2012. — Vol. 489, no. 7415. — P. 295.
17. Opinion evolution in time-varying social influence networks with prejudiced agents / A. V. Proskurnikov [et al.] // IFAC-PapersOnLine. — 2017. — Vol. 50, no. 1. — P. 11896—11901.
18. *Barbera, P.* Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data / P. Barbera // Political Analysis. — 2015. — Vol. 23, no. 1. — P. 76—91.
19. Predicting the political alignment of twitter users / M. D. Conover [et al.] // 2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing. — IEEE. 2011. — P. 192—199.
20. Political polarization on twitter / M. D. Conover [et al.] // Fifth international AAAI conference on weblogs and social media. — 2011.
21. *Volkova, S.* Inferring user political preferences from streaming communications / S. Volkova, G. Coppersmith, B. Van Durme // Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). — 2014. — P. 186—196.
22. *Makazhanov, A.* Predicting political preference of Twitter users / A. Makazhanov, D. Rafiei, M. Waqar // Social Network Analysis and Mining. — 2014. — Vol. 4, no. 1. — P. 193.

23. Beyond binary labels: political ideology prediction of twitter users / D. Preotiuc-Pietro [et al.] // Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). — 2017. — P. 729—740.
24. *Cohen, R.* Classifying political orientation on Twitter: It's not easy! / R. Cohen, D. Ruths // Seventh International AAAI Conference on Weblogs and Social Media. — 2013.
25. *Barbera, P.* How social media reduces mass political polarization. Evidence from Germany, Spain, and the US / P. Barbera // Job Market Paper, New York University. — 2014. — Vol. 46.
26. *Acemoglu, D.* Opinion dynamics and learning in social networks / D. Acemoglu, A. Ozdaglar // Dynamic Games and Applications. — 2011. — Vol. 1, no. 1. — P. 3—49.
27. Opinion dynamics: models, extensions and external effects / A. Sirbu [et al.] // Participatory sensing, opinions and collective awareness. — Springer, 2017. — P. 363—401.
28. *Ceragioli, F.* Consensus and disagreement: The role of quantized behaviors in opinion dynamics / F. Ceragioli, P. Frasca // SIAM Journal on Control and Optimization. — 2018. — Vol. 56, no. 2. — P. 1058—1080.
29. *Clifford, P.* A model for spatial conflict / P. Clifford, A. Sudbury // Biometrika. — 1973. — Vol. 60, no. 3. — P. 581—588.
30. Ergodic theorems for weakly interacting infinite systems and the voter model / R. A. Holley, T. M. Liggett, [et al.] // The annals of probability. — 1975. — Vol. 3, no. 4. — P. 643—663.
31. Network science on belief system dynamics under logic constraints / N. E. Friedkin [et al.] // Science. — 2016. — Vol. 354, no. 6310. — P. 321—326.
32. *Краснощеков, П. С.* Простейшая математическая модель поведения. Психология конформизма / П. С. Краснощеков // Математическое моделирование. — 1998. — Т. 10, № 7. — С. 76—92.
33. *Axelrod, R. P.* The dissemination of culture: A model with local convergence and global polarization / R. P. Axelrod // Journal of conflict resolution. — 1997. — Vol. 41, no. 2. — P. 203—226.

34. *Milgram, S.* Behavioral study of obedience. / S. Milgram // The Journal of abnormal and social psychology. — 1963. — Vol. 67, no. 4. — P. 371.
35. *Brandstatter, H.* Group decision making. Vol. 25 / H. Brandstatter. — Academic Press, 1982.
36. *Bikhchandani, S.* A theory of fads, fashion, custom, and cultural change as informational cascades / S. Bikhchandani, D. Hirshleifer, I. Welch // Journal of political Economy. — 1992. — Vol. 100, no. 5. — P. 992—1026.
37. Social learning and deviant behavior: A specific test of a general theory / R. L. Akers [et al.] // Contemporary Masters in Criminology. — Springer, 1995. — P. 187—214.
38. *Festinger, L.* Social pressures in informal groups; a study of human factors in housing. / L. Festinger, S. Schachter, K. Back. — 1950.
39. *Flache, A.* Small worlds and cultural polarization / A. Flache, M. W. Macy // The Journal of Mathematical Sociology. — 2011. — Vol. 35, no. 1—3. — P. 146—176.
40. Mixing beliefs among interacting agents / G. Deffuant [et al.] // Advances in Complex Systems. — 2000. — Vol. 3, 01n04. — P. 87—98.
41. *Altafini, C.* Consensus problems on networks with antagonistic interactions / C. Altafini // IEEE Transactions on Automatic Control. — 2012. — Vol. 58, no. 4. — P. 935—946.
42. *Бреер, В. В.* Микро- и макромоделли социальных сетей. Ч. 1. Основы теории / В. В. Бреер, Д. А. Новиков, А. Д. Рогаткин // Проблемы управления. — 2014. — № 5.
43. *Mikhailov, A. P.* A Model of Information Warfare in a Society with a Piecewise Constant Function of the Destabilizing Impact / A. P. Mikhailov, A. P. Petrov, O. G. Proncheva // Mathematical Models and Computer Simulations. — 2019. — Vol. 11, no. 2. — P. 190—197.
44. *Mikhailov, A. P.* Models of information warfare / A. P. Mikhailov, N. A. Marevtseva // Mathematical models and computer simulations. — 2012. — Vol. 4, no. 3. — P. 251—259.

45. *Petrov, A. P.* Modeling position selection by individuals during information warfare in society / A. P. Petrov, A. I. Maslov, N. A. Tsaplin // *Mathematical models and computer simulations*. — 2016. — Vol. 8, no. 4. — P. 401—408.
46. *Rashevsky, N.* Studies in mathematical theory of human relations / N. Rashevsky // *Psychometrika*. — 1939. — Vol. 4, no. 3. — P. 221—239.
47. *French Jr, J. R.* A formal theory of social power. / J. R. French Jr // *Psychological review*. — 1956. — Vol. 63, no. 3. — P. 181.
48. *Harary, F.* A criterion for unanimity in French's theory of social power. / F. Harary. — 1959.
49. *DeGroot, M. H.* Reaching a consensus / M. H. DeGroot // *Journal of the American Statistical Association*. — 1974. — Vol. 69, no. 345. — P. 118—121.
50. *Abelson, R. P.* Mathematical models of the distribution of attitudes under controversy / R. P. Abelson // *Contributions to mathematical psychology*. — 1964.
51. *Taylor, M.* Towards a mathematical theory of influence and attitude change / M. Taylor // *Human Relations*. — 1968. — Vol. 21, no. 2. — P. 121—139.
52. *Friedkin, N. E.* A formal theory of reflected appraisals in the evolution of power / N. E. Friedkin // *Administrative Science Quarterly*. — 2011. — Vol. 56, no. 4. — P. 501—529.
53. Opinion dynamics and the evolution of social power in influence networks / P. Jia [et al.] // *SIAM review*. — 2015. — Vol. 57, no. 3. — P. 367—397.
54. *Akers, R. L.* Deviant behavior: A social learning approach / R. L. Akers. — 1973.
55. *Abelson, R. P.* Mathematical models in social psychology / R. P. Abelson // *Advances in experimental social psychology*. Vol. 3. — Elsevier, 1967. — P. 1—54.
56. *Galam, S.* Majority rule, hierarchical structures, and democratic totalitarianism: A statistical approach / S. Galam // *Journal of Mathematical Psychology*. — 1986. — Vol. 30, no. 4. — P. 426—434.
57. The theory of critical phenomena: an introduction to the renormalization group / J. J. Binney [et al.]. — Oxford University Press, 1992.

58. *Xiong, F.* Modeling and predicting opinion formation with trust propagation in online social networks / F. Xiong, Y. Liu, J. Cheng // Communications in Nonlinear Science and Numerical Simulation. — 2017. — Vol. 44. — P. 513—524.
59. *Martins, A. C.* Continuous opinions and discrete actions in opinion dynamics problems / A. C. Martins // International Journal of Modern Physics C. — 2008. — Vol. 19, no. 04. — P. 617—624.
60. Novel multidimensional models of opinion dynamics in social networks / S. E. Parsegov [et al.] // IEEE Transactions on Automatic Control. — 2016. — Vol. 62, no. 5. — P. 2270—2285.
61. *Granovetter, M.* Threshold models of diffusion and collective behavior / M. Granovetter, R. Soong // The Journal of Mathematical Sociology. — 1983. — Vol. 9, no. 3. — P. 165—179. — URL: <https://doi.org/10.1080/0022250X.1983.9989941>.
62. Opinion dynamics and bounded confidence models, analysis, and simulation / R. Hegselmann, U. Krause, [et al.] // Journal of artificial societies and social simulation. — 2002. — Vol. 5, no. 3.
63. Asynchronous opinion dynamics with online and offline interactions in bounded confidence model / Z. Ding [et al.]. — 2017.
64. *Sznajd-Weron, K.* Opinion evolution in closed community / K. Sznajd-Weron, J. Sznajd // International Journal of Modern Physics C. — 2000. — Vol. 11, no. 06. — P. 1157—1165.
65. *Newman, M. E. J.* Networks / M. E. J. Newman. — Oxford university press, 2018.
66. *Barabasi, A.-L.* Emergence of scaling in random networks / A.-L. Barabasi, R. Albert // science. — 1999. — Vol. 286, no. 5439. — P. 509—512.
67. *Watts, D. J.* Collective dynamics of ‘small-world’ networks / D. J. Watts, S. H. Strogatz // nature. — 1998. — Vol. 393, no. 6684. — P. 440.
68. *Hendrickx, J. M.* Open multi-agent systems: Gossiping with random arrivals and departures / J. M. Hendrickx, S. Martin // 2017 IEEE 56th Annual Conference on Decision and Control (CDC). — IEEE. 2017. — P. 763—768.

69. *Franceschelli, M.* Proportional dynamic consensus in open multi-agent systems / M. Franceschelli, P. Frasca // 2018 IEEE Conference on Decision and Control (CDC). — IEEE. 2018. — P. 900—905.
70. *Franceschelli, M.* Stability of Open Multi-Agent Systems and Applications to Dynamic Consensus / M. Franceschelli, P. Frasca // arXiv preprint arXiv:1906.00890. — 2019.
71. *Гантмахер, Ф. Р.* Теория матриц / Ф. Р. Гантмахер. — Физматлит, 2010.
72. *Boyd, S.* Fastest mixing Markov chain on a graph / S. Boyd, P. Diaconis, L. Xiao // SIAM review. — 2004. — Vol. 46, no. 4. — P. 667—689.
73. *Saburov, M.* Reaching a consensus: a discrete nonlinear time-varying case / M. Saburov, K. Saburov // International Journal of Systems Science. — 2016. — Vol. 47, no. 10. — P. 2449—2457.
74. *Friedkin, N. E.* Social influence and opinions / N. E. Friedkin, E. C. Johnsen // Journal of Mathematical Sociology. — 1990. — Vol. 15, no. 3/4. — P. 193—206.
75. *Friedkin, N. E.* The problem of social control and coordination of complex systems in sociology: A look at the community cleavage problem / N. E. Friedkin // IEEE Control Systems Magazine. — 2015. — Vol. 35, no. 3. — P. 40—51.
76. The PageRank citation ranking: Bringing order to the web.Tech. rep. / L. Page [et al.] ; Stanford InfoLab. — 1999.
77. *Brin, S.* The anatomy of a large-scale hypertextual web search engine / S. Brin, L. Page // Computer networks and ISDN systems. — 1998. — Vol. 30, no. 1—7. — P. 107—117.
78. Gossips and prejudices: Ergodic randomized dynamics in social networks / P. Frasca [et al.] // IFAC Proceedings Volumes. — 2013. — Vol. 46, no. 27. — P. 212—219.
79. Ergodic randomized algorithms and dynamics over networks / C. Ravazzi [et al.] // IEEE transactions on control of network systems. — 2014. — Vol. 2, no. 1. — P. 78—87.
80. Distributed randomized algorithms for opinion formation, centrality computation and power systems estimation: A tutorial overview / P. Frasca [et al.] // European journal of control. — 2015. — Vol. 24. — P. 2—13.

81. Learning Political DNA in the Italian Senate / A. Longo [et al.] // 2019 18th European Control Conference (ECC). — IEEE. 2019. — P. 3526—3531.
82. *Wai, H.-T.* Active sensing of social networks / H.-T. Wai, A. Scaglione, A. Leshem // IEEE Transactions on Signal and Information Processing over Networks. — 2016. — Vol. 2, no. 3. — P. 406—419.
83. *Esteban, J.-M.* On the measurement of polarization / J.-M. Esteban, D. Ray // Econometrica: Journal of the Econometric Society. — 1994. — P. 819—851.
84. *Kozitsin, I. V.* Dynamic model of collective decision making / I. V. Kozitsin // VIII Moscow International Conference on Operations Research (ORM2016). — 2016. — C. 138—139.
85. Моделирование политических взглядов российских пользователей социальной сети ВКонтакте / И. В. Козицин [и др.] // Математическое моделирование. — 2019. — Т. 31, № 8. — С. 3—20.
86. *Kozitsin, I. V.* Supplementary information on the paper 'Modelling Russian users' political preferences' / I. V. Kozitsin. — 2019. — URL: <https://doi.org/10.7910/DVN/AKXXUI> ; type: dataset.
87. *Stryker, S.* Symbolic interactionism: A social structural version / S. Stryker. — Benjamin-Cummings Publishing Company, 1980.
88. Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship / K. Garimella [et al.] // Proceedings of the 2018 World Wide Web Conference. — International World Wide Web Conferences Steering Committee. 2018. — P. 913—922.
89. *Frey, D.* Recent research on selective exposure to information / D. Frey // Advances in experimental social psychology. Vol. 19. — Elsevier, 1986. — P. 41—80.
90. Information cascades in feed-based networks of users with limited attention / S. Sreenivasan [et al.] // IEEE Transactions on Network Science and Engineering. — 2016. — Vol. 4, no. 2. — P. 120—128.
91. *Goncalves, B.* Modeling users' activity on twitter networks: Validation of dunbar's number / B. Goncalves, N. Perra, A. Vespignani // PloS one. — 2011. — Vol. 6, no. 8. — e22656.

92. *Вьюгин, В.* Математические основы машинного обучения и прогнозирования / В. Вьюгин. — Litres, 2017.
93. Прикладная статистика: Классификация и снижение размерности / С. А. Айвазян [и др.]. — 1989.
94. *Flach, P.* Machine learning: the art and science of algorithms that make sense of data / P. Flach. — Cambridge University Press, 2012.
95. *Kozitsin, I. V.* Replication data on the papers 'Symmetric Convex Mechanism of Opinion Formation Predicts Directions of Users' Opinions Trajectories' and 'On an Approach to Measure the Level of Polarization of Individuals' Opinions' / I. V. Kozitsin. — 2019. — URL: <https://doi.org/10.7910/DVN/OUZY74> ; type: dataset.
96. *Holme, P.* Nonequilibrium phase transition in the coevolution of networks and opinions / P. Holme, M. E. J. Newman // *Phys. Rev. E.* — 2006. — Nov. — Vol. 74, issue 5.
97. *Kairam, S. R.* The life and death of online groups: Predicting group growth and longevity / S. R. Kairam, D. J. Wang, J. Leskovec // *Proceedings of the fifth ACM international conference on Web search and data mining.* — ACM. 2012. — P. 673—682.
98. Group formation in large social networks: membership, growth, and evolution / L. Backstrom [et al.] // *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining.* — ACM. 2006. — P. 44—54.

Список рисунков

2.1	(а) Динамика мнений для системы из примера 2.3.1. (б) Динамика мнений для системы из примера 2.3.2.	48
2.2	Динамика мнений для $n = 20$, $m = 5$ и $\mu = 0.5$. Начальные мнения агентов откалиброваны в соответствии с нормальным распределением со средним 0.5 и стандартным отклонением 0.25. Траектории мнений агентов-конформистов представлены серым цветом, агентов-неконформистов – черным.	53
2.3	Динамика мнений для той же системы, что представлена на рисунке 2.2 за исключением того, что коэффициенты μ_i агентов-неконформистов калибруются как случайная величина, равномерно распределенная на отрезке $[0, 1]$	54
2.4	Динамика мнений для системы из примера 2.4.1. Параметры системы калиброваны следующим образом. Число агентов $n = 100$. Число конформистов $m = 30$. Коэффициент сопротивления социальному влиянию агентов-неконформистов $\mu = 1/2$. Мнения агентов в начальный момент времени взяты из нормального распределения со средним 0.5 и дисперсией $\sigma^2 = 0.04$. Мнение конформистов $p_{conf}(t)$ есть случайная величина, подчиняющаяся нормальному распределению со средним 0.5 и дисперсией $\sigma^2 = 0.04$. Рассматривается процесс из 50 - ти итераций. Траектории мнений агентов-конформистов показаны серым, остальных агентов - черным. Для такой системы $\rho = 0.7$. К моменту седьмой итерации агенты-неконформисты достигают консенсуса. Если взять $\varepsilon = 0.3$, то для $t \in \{25, 26 \dots, 50\}$ доля итераций, в рамках которых $ p(t) - p_{conf}(t) < \varepsilon$ составляет 0.88. При этом, оценка снизу (2.27) равняется 0.77.	58
3.1	Половозрастная пирамида пользователей Д1.	68
3.2	ПП пользователей Д1.	69
3.3	70
3.4	70

3.5	Типичный вид матрицы ошибок, когда и обучение, и тестирование происходит только на изначальных пользователях Д1.	72
3.6	Скриншот кода, предназначенного для разделения данных Д1 на обучающие и тестовые выборки.	74
3.7	75
3.8	Качество классификации Предсказательной Модели на тестовой выборке. Средняя полнота равна 0.613.	79
3.9	80
3.10	86
3.11	87
3.12	Гистограмма распределения степеней вершин графа G_u . Оси абсцисс и ординат имеют логарифмический масштаб. Приблизительная прямолинейная форма средней и хвостовой частей гистограммы (начиная с $k_{min} = 15$) говорит о том, что распределение степеней вершин соответствует степенному закону. . .	88
3.13	89
3.14	89
3.15	92
3.16	93
3.17	94
3.18	95
3.19	96

Список таблиц

1	Распределение пользователей Д1 по классам	64
2	Надежность признаков для пользователей Д1	69
3	Влияние порога аномалий th на качество классификации Предсказательной Модели	75
4	Сравнение результатов классификации Д2 с помощью Предсказательной Модели с результатами выборов Президента РФ 2018 года.	81
5	Ранжировка городов пользователей Д2 с учетом уровня поддержки В.В. Путина.	82
6	Динамика изменения среднего, стандартного отклонения σ мнений пользователей U , а также ассортативности и индекса Джини.	85
7	Разбиения пользователей в соответствии с магнитудами изменений оценок их мнений	118

Приложение А

Численности групп пользователей

Таблица 7 — Разбиения пользователей в соответствии с магнитудами изменений оценок их мнений

$ \tilde{p}_i(t_{k+1}) - \tilde{p}_i(t_k) \in$	Первая итерация	Вторая итерация
$(0, 0.1]$	1372526	1380424
$(0.1, 0.2]$	40301	18649
$(0.2, 0.3]$	9291	4329
$(0.3, 0.4]$	2258	1364
$(0.4, 0.5]$	894	535
$(0.5, 0.6]$	363	200
$(0.6, 0.7]$	149	101
$(0.7, 0.8]$	74	67
$(0.8, 0.9]$	37	15
$(0.9, 1]$	8	6