



**Politecnico
di Torino**

Dipartimento
di Elettronica
e Telecomunicazioni

ANTON PROSKURNIKOV
Associate Professor

ОТЗЫВ

члена диссертационного совета Проскурникова Антона Викторовича на диссертацию **Чжао Чи (Zhao Chi)** на тему «*Моделирование динамики бинарных мнений в социальных сетях сложных конфигураций*», представленную на соискание ученой степени

кандидата технических наук по научной специальности

2.3.1. Системный анализ, управление и обработка информации, статистика.

Моделирование динамики мнений – новая область научного знания, находящаяся на стыке социальных наук, с одной стороны, и теории динамических систем – с другой. Хотя первые работы по моделированию динамики мнений появились еще в 1950-70-е гг., бурный интерес к данной тематике возник именно в последние десятилетия с появлением социальных сетей и возросшей ролью онлайн-взаимодействий в социальных процессах.

В диссертации предлагается учитывать двухуровневую природу социальных коммуникаций, а именно, различие между публичными высказываниями и частными убеждениями (два слоя в сети коммуникаций). Такой подход отражает реальные аспекты социального взаимодействия, наблюдаемые в обществе, когда публичные и частные мнения часто не совпадают. Помимо собственно динамических моделей, диссертация затрагивает и ряд других важных тем, в частности, исследует меры центральности вершин в двуслойных графах. При этом автор предлагает новый метод на основе теоретико-игрового подхода для оценки важности признаков, участвующих в построении сложных моделей искусственного интеллекта. В целом, работу можно охарактеризовать как одну из первых (по моим сведениям) диссертацию по *теории сетей* (network science), выполненную в Санкт-Петербургском государственном университете. Особую ценность исследованию придает его междисциплинарный характер, объединяющий методы теории игр, теории графов и машинного обучения.

Объем и структура работы. Диссертация состоит из введения, четырех глав, заключения, списка цитированной литературы и приложения. Во введении дается описание структуры диссертации, основных результатов и текущего состояния науки в данной области. **Глава 1** посвящена разработке макроскопической модели динамики мнений в двухслойных социальных сетях. Автор предлагает обобщение классической модели скрытого избирателя (CVM), вводя возможность взаимодействия между агентами не только на публичном, но и на частном уровне. Эта модификация, названная общей моделью скрытого избирателя (GCVM), более адекватно отражает реальные социальные процессы, где люди публично могут выражать мнения, отличные от их внутренних. В главе исследуется влияние различных структур сети (полные, циклические, звездообразные и другие конфигурации) на процесс достижения консенсуса. Особое внимание уделяется анализу времени достижения консенсуса и вероятности победы одного мнения из двух возможных в зависимости от начальных условий и параметров модели. **Глава 2** развивает подходы первой главы, переходя от



макроскопического описания к микроскопической модели, основанной на реальной сетевой структуре. Здесь формализуется понятие двухслойной сети с узлами-репликами. Автор выявляет зависимость между временем достижения консенсуса и некоторыми характеристиками сети, такими как длина кратчайшего пути между агентами. Оценивается влияние отдельных параметров (например, структуры сети) на динамику мнений. Дополнительно исследуется влияние комбинаций параметров с использованием методов кластеризации. В **главе 3** делается акцент на применение разработанных моделей к реальным социальным сетям на примере «клуба каратэ» Закари. Основное внимание уделяется связи между топологическими характеристиками сети (меры центральности) и динамикой мнений. Автор предлагает метод упрощения двухслойной сети до однослойной взвешенной, что позволяет применять классические подходы к определению центральности вершин. Разрабатываются эффективные алгоритмы вычисления теоретико-игровых мер центральности, демонстрируется их связь с временем достижения консенсуса. Результаты показывают значимую корреляцию между центральностью влиятельных узлов и временем достижения консенсуса. **Глава 4** переносит полученные результаты в область объяснимого искусственного интеллекта. На основе мер центральности разрабатывается новый метод оценки важности признаков в моделях машинного обучения - ShapG. Этот метод, использующий теоретико-графовое представление взаимосвязей между признаками, демонстрирует преимущества по точности и эффективности по сравнению с существующими подходами (SHAP, LIME). Практическая ценность метода иллюстрируется на реальных наборах данных (два датасета), показывая его применимость для задач интерпретации сложных моделей.

Апробация результатов подтверждена выступлениями автора на нескольких международных и российских профильных конференциях.

Публикации автора насчитывают четыре статьи, все они проиндексированы в Scopus, и две программы, одна имеет свидетельство регистрации, а другая – выложена в общий доступ в сети Интернет.

В работе получены новые оригинальные результаты, определяющие новизну диссертации, среди которых можно отметить следующие:

1. Разработана общая модель скрытого избирателя (GCVM) - модификация классической модели CVM, учитывающая взаимодействия агентов как на публичном, так и на частном уровнях социальной сети.
2. Экспериментально установлены закономерности влияния структуры сети на динамику мнений: выявлена зависимость времени достижения консенсуса от конфигурации сети, обнаружены условия, при которых внутренние взаимодействия не влияют на достижения консенсуса.
3. Предложен метод преобразования двухслойных сетей с заданной на ней динамикой мнений в однослойные взвешенные графы, позволяющий применять классические методы сетевого анализа однослойных сетей к сложным двухуровневым структурам.
4. Разработаны новые алгоритмы для эффективного вычисления теоретико-игровых мер центральности в сложных сетях, обладающие высокой точностью



**Politecnico
di Torino**

Dipartimento
di Elettronica
e Telecomunicazioni

ANTON PROSKURNIKOV
Associate Professor

- вычислений, существенно меньшей вычислительной сложностью по сравнению с существующими аналогами.
5. Установлены статистически значимые корреляции между мерами центральности узлов и временем достижения консенсуса, плотностью внутреннего графа и динамикой мнений.
 6. Создан новый метод объяснимого ИИ (XAI) - ShapG, обладающий более высокой точностью оценки важности признаков по сравнению с SHAP и LIME, меньшей вычислительной сложностью, возможностью глобальной интерпретации моделей. Преимуществом разработанного алгоритма является его открытая версия, выложенная в виде пакета для Python.
 7. Разработано программное обеспечение для моделирования динамики мнений в различных сетевых структурах, вычисления мер центральности в сложных сетях.

Проведенное исследование обладает существенной **практической ценностью** для различных областей. Разработанные модели позволяют моделировать процессы формирования общественного мнения, выявлять ключевых агентов влияния и разрабатывать эффективные стратегии информационного воздействия. Такие модели могут найти применение в политике и маркетинге, где требуется разрабатывать кампании по продвижению идей и продуктов и прогнозировать реакцию целевой аудитории. Особую практическую значимость представляет разработанный метод ShapG, который вносит вклад в развитие объяснимого искусственного интеллекта. Этот метод существенно улучшает интерпретацию сложных моделей машинного обучения, что особенно востребовано в прикладных областях, например, в медицине, где определение важных факторов в предсказании исхода лечения пациентов или эффективности лечения тем или иным препаратом имеет критическое значение. При этом предложенные оптимизированные алгоритмы не только повышают точность анализа, но и способствуют снижению вычислительных затрат, что соответствует современным требованиям энергоэффективности и экологичности ИТ-решений.

Обоснованность и достоверность основных результатов и выводов диссертации Чжао Чи обеспечивается корректностью постановок задач работы, проведенными экспериментами, глубокому сравнительному анализу работы предложенных методов с существующими, а также публикациями в журналах и выступлениями на известных научных конференциях.

Несмотря на положительное впечатление, к автору имеются **вопросы и замечания**:

1. Целый ряд фактических неточностей имеется в историческом обзоре (раздел Введение, с.7-9), который не имеет прямого отношения к остальной диссертации и, по моему мнению, охватывает слишком широкий круг вопросов и смешивает совершенно разные классы моделей, которые имеют свои особенности.



**Politecnico
di Torino**

Dipartimento
di Elettronica
e Telecomunicazioni

ANTON PROSKURNIKOV
Associate Professor

Винер действительно обозначил важность кибернетики в общественных науках, но никогда не использовал термин «социокибернетика», этот термин введен существенно позже известным нидерландским социологом Феликсом Гейером (Felix Geyer) в серии работ, опубликованных в 1990-е годы и позже.

Работы Изинга начала 20го века по ферромагнитам вообще к динамике мнений отношения не имели, хотя введенные им переменные и послужили мотивацией для специалистов по статистической физике, в частности, супругов Шнайда и Шнайд-Верон, предложивших свою модель («модель Шнайда») в 2000 году.

На странице 7 модель голосующего отнесена к «макроскопическим», но при ее описании упоминаются агенты и их взаимодействия с соседями, что характерно как раз для *микроскопических* моделей. На самом деле, есть как макроскопические, так и микроскопические версии описываемых на с.7 моделей.

Модель ДеГроота была предложена не для описания динамики мнений, а как процедура последовательного согласования численных оценок в группе экспертов, которым нужно прийти к консенсусу. Упомянув динамику мнений, скорее следовало бы сослаться на более ранние работы Френча и Харари 1950-х годов. Далее сказано, что в этой модели «каждый индивид обновляет своё мнение на основе своего собственного мнения и мнений соседей», но это, вообще говоря, относится к абсолютно всем динамическим моделям на графах, а не только к модели ДеГроота и ее обобщениям. Другое дело, что в этой модели происходит вычисление взвешенного среднего мнений по всем соседям.

Фамилия Джонсен пишется через «е» (Johnsen). Модель Фридкина-Джонсена характеризуется не просто упрямыми агентами (они могут быть добавлены к любой модели динамики мнений, в том числе и Де Гроота), а наличием параметра восприимчивости к общественному мнению для всех агентов – от 0 (упрямый агент) до 1 (агент, подчиняющийся модели ДеГроота). Это позволяет, сохраняя линейность модели, получать очень широкий спектр финальных мнений.

Фамилия Weisbuch произносится Вайсбух, а не Вейбуш.

На с.9 работа [9] процитирована в связи с *колебаниями* мнений. Это достаточно странно, поскольку в модели Хегсельманна-Краузе мнения всегда сходятся. Для разнородных агентов доказательство этого до сих пор, по-видимому, является открытой проблемой, но ни одного примера колеблющихся мнений в литературе нет, а в работе [9] мне не удалось даже найти самого слова oscillation (колебание).

Исходная работа Краснощекова не содержала динамических моделей, его так называемые модели поведения – это системы алгебраических уравнений, хотя и



мотивированные теорией марковских цепей. Динамическая модель на основе теории Краснощекова была введена лишь в недавних работах Козицина.

Вообще не упомянуты модели с бинарными мнениями и мажоритарным принципом принятия решений, в которых агент меняет свое мнение, если оно отличается от большинства его соседей. Например, есть пороговые модели Грановеттера (M. Granovetter) и мажоритарные модели Галама (S. Galam). Прямое отношение к теме диссертации имеет модель ЕРО (expressed and private opinion), введенная в работе [Ye, M., Qin, Y., Govaert, A., Anderson, B. D. O., & Cao, M. (2019). An influence network model to study discrepancies in expressed and private opinions. *Automatica*, 107 (9), 371-381] как раз для описания различий между высказываемыми (публичными, в терминологии соискателя) и частными мнениями и предлагающая объяснения знаменитых экспериментов Соломона Эша. Данную работу (и ряд других статей по ЕРО) точно следовало упомянуть.

На с.10 смешиваются понятия социальной власти и меры центральности. Действительно, в некоторых моделях (ДеГроота, Фридкина-Джонсена) влияние агента на финальное мнение группы можно охарактеризовать некоторой мерой центральности, которую часто называют social power или influence centrality. Но на самом деле центральность (то есть некоторый ранг вершины, вычисляемый по определенному алгоритму) не обязательно характеризует степень влияния соответствующего агента в системе, да и вообще определена для любого графа.

2. В жизни наше мнение бинарно не всегда. Например, наше мнение о прочитанной книге или каком-то блюде в ресторане может иметь больше двух значений – в простейшем случае «хорошо, плохо, средне». Известны ли автору модели с такими мнениями? Есть ли возможность развития представленных им моделей на троичные или произвольные дискретные мнения? Или на многомерные мнения, где компоненты (пусть даже и бинарные) выражают отношения к разным объектам или вопросам и могут быть логически связаны (мое мнение о качестве еды в ресторане, мнение о владельце ресторана, мнение о том, уместно ли туда пригласить друзей для отмечания праздника)?
3. Мера центральности – это всего лишь алгоритм, позволяющий упорядочить узлы графа. Самые простые – центральность собственного вектора (eigenvector centrality), центральность Каца (Katz centrality), PageRank. Была ли попытка проверить корреляцию между этими более простыми мерами центральности и временем консенсуса? Кроме того, меры центральности, например уровень посредничества (betweenness) существуют не только для вершин, но и для ребер. Были ли попытки развития таких мер центральности? Какие ребра наиболее существенны для достижения консенсуса?



**Politecnico
di Torino**

Dipartimento
di Elettronica
e Telecomunicazioni

ANTON PROSKURNIKOV
Associate Professor

4. Понятно, что в теории ИИ многие понятия сложно определить математически. Тем не менее, на мой взгляд, понятие объяснимости следовало бы как-то формализовать, если не математически, то хотя бы словесно. Например, относится ли физически информированное обучение (physics-informed ML) к объяснимому ИИ? Более того, на с.120 автор утверждает, что определенная операция над графом *улучшает объяснимость*. В чем это проявляется? Можно ли это утверждение облечь в математическую форму, дать критерии как сравнивать объяснимость для разных наборов данных и разных методов? Далее, на с.137 сказано, что метод *теоретически применим к любой модели*. Речь идет все-таки об определенном классе задач, например задач классификации или машинном обучении вообще? Под моделями понимается вообще все что угодно, включая системы управления с внутренними переменными состояниями и сложной динамикой -- диффузионные модели, резервуары, импульсные (spiking) нейронные сети, нейронные обыкновенные и стохастические дифференциальные уравнения (neural ODE/SDE)? Мне кажется, что следует точно характеризовать область применимости, поскольку понятия «объяснимости» для каждого класса моделей должно отличаться и включать в себя разные требования.

Сделанные выше замечания и сформулированные вопросы не снижают научной ценности работы и не влияют на общее положительное впечатление от диссертационной работы. По актуальности темы, объему проделанной работы и достоверности представленных результатов, значимости выводов работа полностью соответствует уровню кандидатской диссертации по техническим наукам по научной специальности 2.3.1. Системный анализ, управление и обработка информации, статистика.

Заключение. Диссертация Чжао Чи на тему «Моделирование динамики бинарных мнений в социальных сетях сложных конфигураций» соответствует основным требованиям, установленным Приказом от 19.11.2021 №11181/1 «О порядке присуждения ученых степеней в Санкт-Петербургском государственном университете», соискатель Чжао Чи заслуживает присуждения ученой степени кандидата технических наук по специальности 2.3.1. Системный анализ, управление и обработка информации, статистика. Нарушения пунктов 9 и 11 указанного Порядка в диссертации мной выявлены не были.

член диссертационного совета
доктор физико-математических наук,
Туринский политехнический университета
(Polytechnic University of Turin), Италия

09/04/2025


Проскурников Антон Викторович