

Федеральное государственное автономное  
образовательное учреждение высшего образования  
«Национальный исследовательский университет  
«Высшая школа экономики»

*на правах рукописи*

Киселёв Дмитрий Андреевич  
**Рекомендательные системы,  
основанные на графах,  
с использованием непрерывных представлений  
сетей**

РЕЗЮМЕ

диссертации на соискание ученой степени  
кандидата компьютерных наук

Москва – 2022

Диссертационная работа выполнена в федеральном государственном автономном образовательном учреждении высшего образования «Национальный исследовательский университет «Высшая школа экономики».

Научный руководитель: Макаров Илья Андреевич, Ph.D., старший научный сотрудник, АНО «Институт искусственного интеллекта»; Доцент, Факультет компьютерных наук, старший научный сотрудник, Лаборатория алгоритмов и технологий анализа сетевых структур, НИУ ВШЭ.

# 1 Тема диссертации

В современном мире рекомендательные системы являются одной из ключевых компонент бизнеса разных компаний от электронной коммерции до социальных медиа [2, 10]. Рекомендательные системы помогают пользователю ориентироваться в большом объеме различных товаров, услуг и контента (для простоты назовем все это объектами), позволяя улучшить пользовательский опыт. Методы рекомендательных систем варьируются от классических техник дополнения матриц (matrix completion) до современных сложных моделей на последовательностях (sequential recommenders), вдохновленных моделями из области обработки естественного языка. Одним из многообещающих подходов является представление задачи рекомендаций в виде предсказания связей (link prediction) на графе взаимодействий пользователей и объектов. Для решения этой задачи активно используются методы машинного обучения на графах. Эти методы быстро развиваются последние несколько лет, но все еще нуждаются в доработке и адаптации для эффективного применения к задаче рекомендаций.

## 2 Объект исследования

Опиши здесь задачу рекомендаций + вставь generic картинку с процессом рекомендаций

Объект исследования данной работы являются рекомендательные системы. Рисунок 1 описывает обобщенную схему решения задачи рекомендаций. На вход, модель получает набор пользователей, объектов, их признаки, истории взаимодействий. На выходе модель для подбирает для пользователя наиболее релевантные объекты, которые оптимизируют показатели вовлеченности (пр.: время проведенное в сервисе) или ожидаемый бизнес результат (пр.: ожидаемую прибыль с рекламного показа баннера). В секции 3.2 более подробно описаны классические подходы к решению рекомендательных систем и их проблемы.

## 3 Субъект исследования

Недавно было предложено рассматривать задачу рекомендаций как задачу на графах [48]. Основная идея – заменить матрицу взаимодействий пользователей и объектов аналогичным (двудольным) графом с гетерогенными ребрами (рисунок 2). Такой подход позволяет совместить коллаборативные и контентные модели. Более того, графовое представление задачи позволяет находить

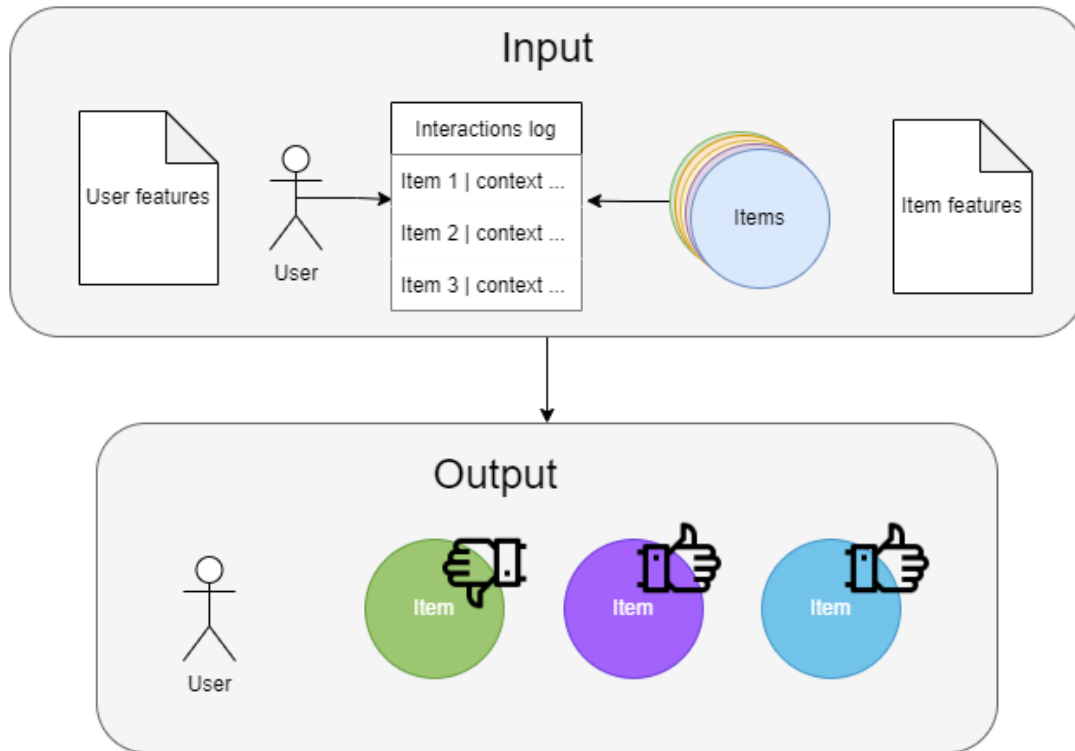


Рис. 1: Обобщенная схема работы моделей рекомендательных систем

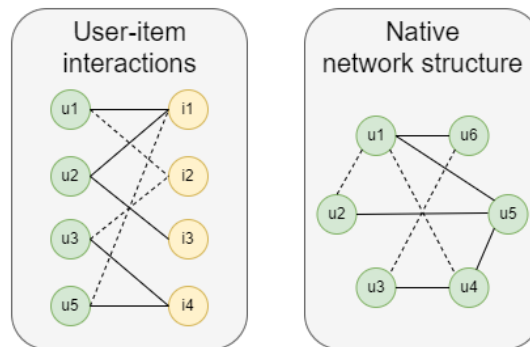


Рис. 2: Представление графа взаимодействия пользователей и объектов для задачи рекомендаций

похожих пользователей и объекты, благодаря простому обходу графа. Поэтому использование топологии графа играет важную роль в улучшении качества рекомендаций. С использованием графа задача дополнения матрицы превращается в задачу предсказания будущих связей на графе (на рисунке 2 хотим

понять появятся ли ребра, изображенные пунктиром). Обычно, для решения такой задачи используют такие, уже хорошо изученные проблемы машинного обучения, как ранжирование и бинарная классификация на присутствие ребра в графе взаимодействий.

Рисунок 3 описывает процесс решения задачи предсказания связей на графе. На вход получаем граф пользовательский взаимодействий. Далее, необходимо закодировать его структуру и другие необходимые свойства (подробнее в секции 3.3). Стратегии векторизации можно разделить на три условных блока: матричные факторизации, подход на основе случайного блуждания, графовые нейронные сети. Первая группа методов берет одно из табличных представлений графа и снижает его размерность, чтобы получить вектора. Вторая группа методов семплирует случайные блуждания и оптимизирует вероятность встречаемости разных вершин в таких блужданиях методом Skip-gram [23]. Последнюю группу методов можно обобщить в рамках фреймворка передачи сообщений (message-passing). В данном фреймворке мы сначала применяем трансформацию и нелинейность к входным признакам вершин, помещая полученный вектор в виде сообщения на ребро. Вторым шагом мы агрегируем все сообщения для конкретной вершины. После получения непрерывных представлений графа мы можем решать последующую задачу предсказания связей используя стандартные подходы к классификации табличных данных. Дополнительно можно также сохранять специфичные свойства графов или предобучать модели используя задачи самообучения.

В рамках диссертации разносторонне изучены различные аспекты непрерывных представлений графов, их свойств и как они влияют на задачу предсказания связей. Более подробно цели и задачи работы описаны в секции 3.3.

### **3.1 Актуальность темы исследования**

За последние 5-10 лет поведение пользователей в онлайн сервисах значительно изменилось, ускорилась динамика изменений в поведении пользователей. Она приводит к тому, что классические подходы становится сложнее применять с пользой для клиента и бизнеса [45, 31]. Поэтому необходимы новые подходы, которые позволяют лучше решать проблемы, к которым приводит изменение поведения пользователей и появление нового контента. Графовое представление задачи рекомендаций позволяет совместить лучшие части разных подходов для более качественного решения задачи рекомендаций.

### **3.2 Проблемы моделей рекомендательных систем**

Рисунок 4 описывает задачу рекомендаций в динамической среде. На входе в задачу рекомендаций мы имеем наборы пользователей и товаров, их истории

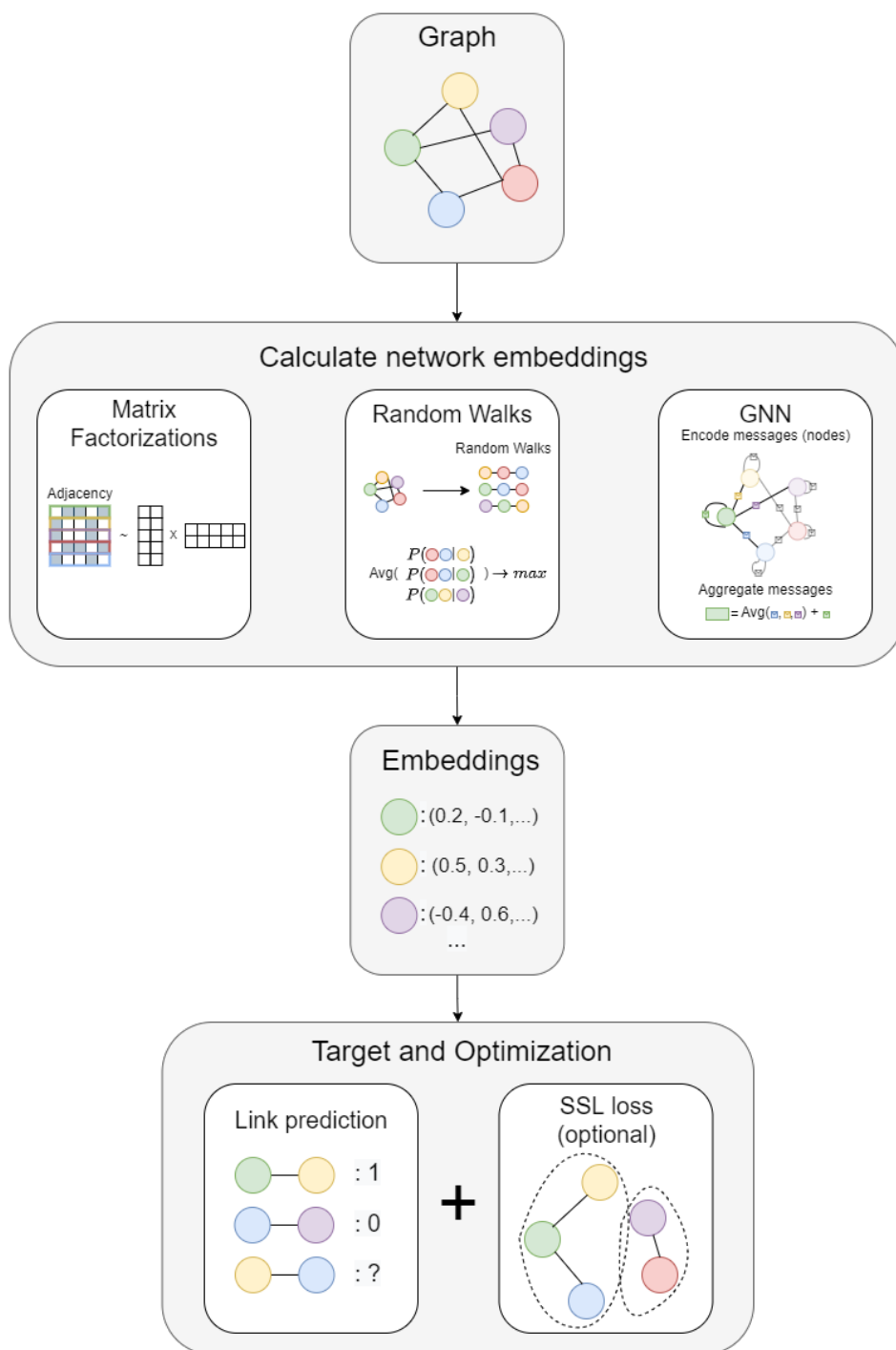


Рис. 3: Процесс предсказания связей в задаче рекомендательных систем

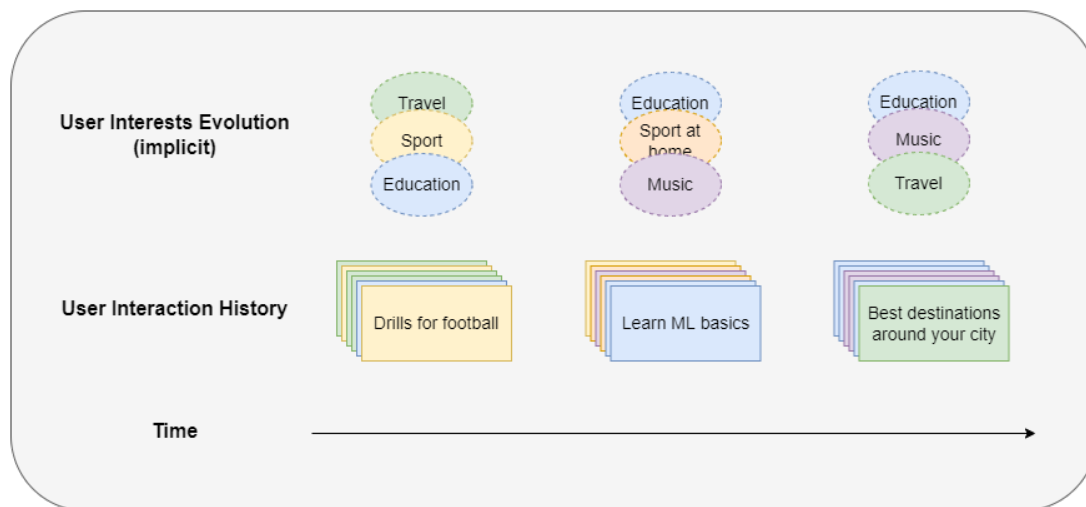


Рис. 4: Основные теоретические проблемы построения рекомендательной системы на основе графов, вызванные динамикой среды

взаимодействий. Наша задача подобрать такой контент, который будет интересен для пользователя в будущем. Допустим у нас есть три периода времени (1) до пандемии, (2) пандемия и жесткие ограничения, (3) небольшие ослабления ограничений. Предпочтения пользователя, которые нам в явном виде неизвестны, меняются со временем по двум причинам: изменилась ситуация в мире и пользователь просто не может больше путешествовать и ходить в спортзал, его собственные интересы меняются. На данном примере мы можем разобрать ряд проблем, которые возникают из-за динамики. Во-первых, у нас появляется новый контент, с которым никакие пользователи еще не взаимодействовали (холодный старт) [38]. Более того, из-за изменений поведения пользователя нам становится непонятно, как именно изменились его интересы (изменение распределения в данных) [5] и нужно ли нам рекомендовать что-то значительно отличающееся от того, что он любил раньше (петля обратной связи) [21]. Наконец, в данных представлена временная структура [45]: в данном примере со временем пользователю начинает больше нравиться музыка и перестает привлекать спорт.

В отдельности, есть разные подходы, которые решают вышеупомянутые проблемы, но в большинстве случаев они остаются подвержены другим. Наиболее популярными моделями рекомендательных систем являются матричные факторизации [42]. Основная идея таких моделей – построить латентные представления пользователей и объектов. Простое перемножение таких факторов позволяет предсказать рейтинг, который пользователь потенциально поставит объекту. Однако, матричные факторизации являются статичными моделями,

они не учитывают признаки пользователей и объектов, не позволяют работать с новыми сущностями. То есть, они сталкиваются со всем спектром упомянутых проблем. Факторизационные машины [34] обобщают матричные факторизации и позволяют учитывать характеристики товаров, тем самым решая проблему холодного старта, но при этом не могут учитывать коллаборативную информацию для новых пользователей и объектов до полного обновления модели. Контентные модели [32] пытаются подобрать наиболее близкий объект к просмотренным по его внутренним характеристикам (текст, картинка). Такие модели обычно не рассматривают взаимодействия пользователей и товаров, поэтому также не способны решить проблему динамики в действиях пользователей. Для того, чтобы учитывать динамику, часто ищут похожие объекты только на те объекты, с которыми пользователь взаимодействовал последними. Для более серьезного учета динамики и последних изменений в выборе пользователя были предложены модели на последовательностях [45]. Они рассматривают пользователя как последовательность объектов и применяют техники из обработки естественного языка для получения финального представления [53, 39]. В качестве базовых представлений объектов могут использоваться контентные признаки. Поэтому такие модели справляются с холодным стартом для новых объектов. Тем не менее, из-за представления пользователя в виде последовательности объектов его признаки, обычно, уходят из рассмотрения, поэтому такие рекомендации все равно дают слабые начальные рекомендации до сбора первой обратной связи. Таким образом, различные типы рекомендательных систем используются для решения разных проблем, но не способны решить все сразу, поэтому необходим новый взгляд на эту задачу, который имеет потенциал для объединения вышеупомянутых типов моделей.

### 3.3 Цели и задачи работы

На картинке 5 описаны основные задачи, которые необходимо решить для того, чтобы создать эффективную рекомендательную систему решающую упомянутые проблемы рекомендательных систем: холодный старт, учет временной структуры, петли обратной связи и изменения распределения данных.

Как уже было упомянуто, задача рекомендаций на графах сводится к бинарной классификации пар вершин на факт их существования. Такие подходы требуют вещественные представления дискретных подструктур графа (вершин и ребер). Поэтому **первой задачей** является анализ существующих методов и подходов для построения непрерывных представлений графов. Однако, простое использование структуры графа позволяет нам давать рекомендации только для пользователей и объектов, которые уже имеют связи в графе. Поэтому **следующими шагами** являются учет контента вершин и ребер, их эффективное смешивание с коллаборативной информацией. **Четвер-**



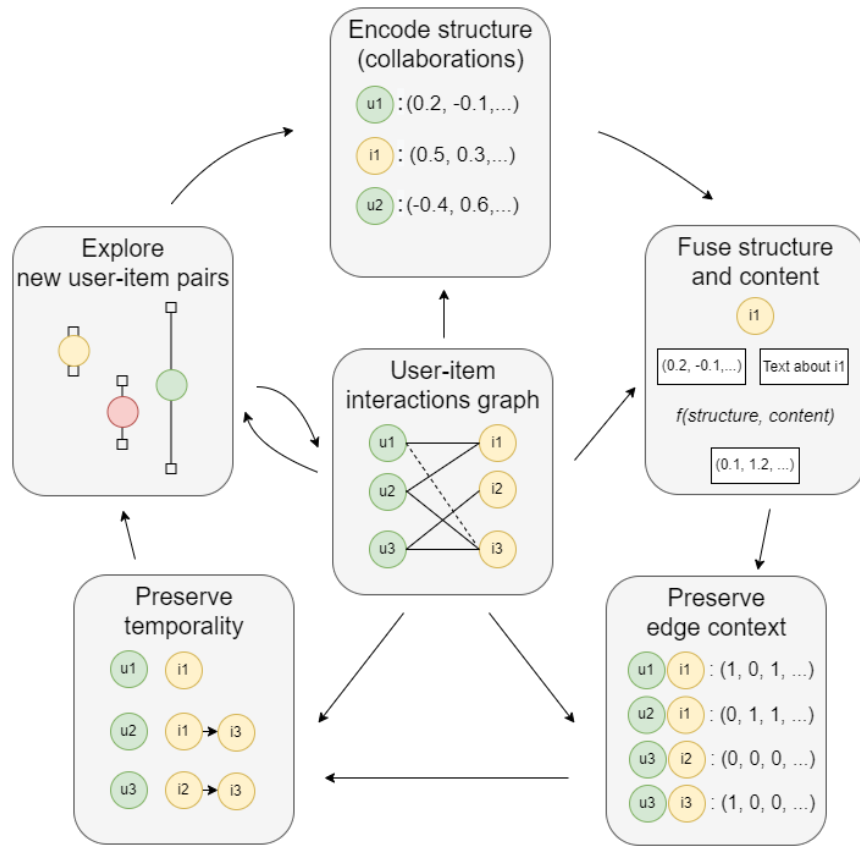


Рис. 5: Структура исследования. Мы постепенно усложняем рекомендательную систему, добавляя новые модель и методы учета различных свойств графов. Сначала мы разбираем сценарий со статичным графов рекомендаций и применяем непрерывные графовые представления для разработки простой коллаборативной рекомендательной системы (a). Далее, мы добавляем атрибуты вершин (b) и ребер (c), чтобы построить гибридную рекомендательную систему, способную решить проблему холодного старта. Наконец, мы рассматриваем динамику графа (d) и методы исследования неизвестных пар пользователь-объект (e), чтобы позволить рекомендательной системе адаптироваться к изменениям в поведении и окружающей среде. Каждый компонент позволяет построить следующий и улучшает итоговую систему, позволяя построить сильную отправную точку для графовый рекомендательных систем, решающих проблемы холодного старта, темпоральность, сдвигов в распределениях данных и петель обратной связи более эффективно.

**тый шаг** – учет временной структуры графа, мы хотим учитывать эволюцию интересов пользователя, учитывать каузальность его действий. Для решения

проблем изменения распределения данных и петли обратной связи необходимо исследовать новые возможные взаимодействия пользователя и объектов. Поэтому **последним шагом** является исследование таких новых состояний. После этого модель дает рекомендацию пользователю, после получения реакции, мы можем обновить граф взаимодействий, а также непрерывные представления вершин.

Таким образом мы выделили пять задач. В рамках диссертации мы рассматриваем их изолировано друг от друга и последовательно усложняем модели для получения итогового варианта, который решает все приведенные задачи. В основном блоке диссертации всего пять глав (статей), каждая из которых посвящена отдельной задаче и ее решению. Более четко, задачи могут сформулированы следующим образом

1. Сравнить методы непрерывных представлений сетей чтобы найти оптимальные применения к разным типам графов и атрибутов в рамках единого подхода
2. Предложить эффективную стратегию для использования как параметров **вершин**, так и структуры графа применительно к задаче предсказания связей (рекомендаций)
3. Предложить эффективную стратегию для использования как параметров **ребер**, так и структуры графа применительно к задаче предсказания связей (рекомендаций)
4. Предложить подход для эффективного извлечения и сохранения темпоральной информации графов
5. Предложить подход для исследования неизвестных возможных взаимодействий пользователей и объектов, чтобы обрабатывать проблемы сдвигов распределений данных и петель обратной связи.

**Основной целью** работы является создать новые графовые методы и показать, что они наилучшим образом подходят для решения проблем динамики среды, интенсивной динамики поведения пользователей в задаче рекомендательных систем.

## 4 Основные положения, выносимые на защиту

Данная секция описывает основной вклад достигнутый представленной диссертацией, его новизну, теоретическую и практическую значимость, исследовательскую методологию и обоснованность полученных результатов.

### Основные положения, выносимые на защиту:

1. Детальная таксономия подходов к построению непрерывных представлений графов и их применений [17]
2. Оценка ведущих подходов к построению непрерывных представлений графов для задачи предсказания связей [17]
3. Методология автоматического извлечения признаков для графов с текстовыми атрибутами вершин в задаче предсказания связей [19]
4. Новый подход к совместному построению представлений вершин и ребер, основанный на самообучении, использующий двойственное представление графа [18]
5. Новый подход к построению темпоральных непрерывных представлений графов, показывающий ведущие результаты в различных темпоральных задачах графового машинного обучения [20]
6. Стандартизированная среда для оценки качества темпоральных непрерывных представлений графов и сравнение ведущих моделей в рамках единого подхода к тренировке и валидации, дающий новое понимание и проясняющий реальное качество моделей в сравнении с тем, что описано в оригинальных статьях [20]
7. Новая стратегия исследования неизвестных состояний (exploration) на основе персонализированного пейджеранка [26] (Personalized PageRank) [16]
8. Новая стратегия исследования неизвестных состояний на основе идеи маленького мира [26] (Personalized PageRank) [16]
9. Методология применения предложенных методов к задаче рекомендательных систем с онлайн адаптацией в динамичных средах [16]

**Научная новизна.** Данная работа концентрируется на молодой, пока слабо изученной задаче графовых рекомендательных систем и использования в ней методов автоматического построения признаков. Диссертация изучает пять основных компонент необходимых для успешного построения рекомендаций на графах: выделение и кодирование структурных признаков, эффективное смешение структурных признаков с атрибутами вершин, построение экспрессивных представлений ребер с учетом контекстной информации, учет динамики графа и исследование малоизученных пар пользователь-объект. Каждой из задач посвящена отдельная работа, в каждой из которых проведены обзор, сравнение и анализ существующих методов, предлагаются стратегии их оптимального использования. Для последних трех задач (представления ребер с учетом контекста, динамика графа и исследование новых состояний) предложены новые модели и методы, которые улучшают качество на задачах предсказания связей и рекомендаций.

**Теоретическая и практическая значимость.** Первые две статьи делают выводы о текущем прогрессе в области учета структурной информации графа и эффективном ее объединении с признаками вершин, предлагают новые направления и взгляд на будущее развитие области. Также, они предлагают советы для практического применения непрерывных представлений графов. Последние три статьи также предлагают новые модели, которые улучшают качество в прикладных данных. Кроме того, четвертая статья представляет стандартизированный подход к оценке качества, который позволяет ускорить исследования в области темпоральных представлений графов и рекомендательных систем, основанных на графах.

**Методология исследования.** Работа основана на теории геометрического глубокого обучения, графовых нейронных сетях, машинном обучении на графах, науки о сетях, рекомендательных системах, классическом машинном обучении и статистике.

**Обоснованность полученных результатов.** Обоснованность результатов достигается за счет комплексных и исчерпывающих экспериментов, включая расчеты доверительных интервалов метрик качества и сравнение с другими ведущими моделями и методами.

**Финансирование.** Исследования проведены при поддержке Факультета Компьютерных Наук, НИУ ВШЭ; Российского Научного Фонда; Программы фундаментальных исследований НИУ ВШЭ; вычислительных ресурсов отдела суперкомпьютерного моделирования НИУ ВШЭ.

## 5 Публикации и апробация работы

### 5.1 Публикации повышенного уровня

1. Ilya Makarov, Dmitrii Kiselev, Nikita Nikitinsky, and Lovro Subelj. Survey on graph embeddings and their applications to machine learning problems on graphs. *PeerJ Computer Science*, 7:e357, 2021 [17]
2. Ilya Makarov, Mikhail Makarov, and Dmitrii Kiselev. Fusion of text and graph information for machine learning problems on networks. *PeerJ Computer Science*, 7, 2021 [19]
3. Ilya Makarov, Ksenia Korovina, and Dmitrii Kiselev. JONNEE: Joint network nodes and edges embedding. *IEEE Access*, 2021 [18]
4. Ilya Makarov, Andrey Savchenko, Arseny Korovko, Leonid Sherstyuk, Nikita Severin, Dmitrii Kiselev, Aleksandr Mikheev, and Dmitrii Babaev. Temporal network embedding framework with causal anonymous walks representations. *PeerJ Computer Science*, 8:e858, 2022 [20]
5. Dmitrii Kiselev and Ilya Makarov. Exploration in sequential recommender systems via graph representations (on review). *IEEE Access*, 2022 [16]

### 5.2 Прочие публикации

Dmitrii Kiselev and Ilya Makarov. Prediction of new itinerary markets for airlines via network embedding. In *International Conference on Analysis of Images, Social Networks and Texts*, pages 315–325. Springer, 2019 [15]

### 5.3 Доклады на конференциях и семинарах

1. Internal seminar of the Sber Recommender Systems Platform, 1 June 2022. Topic: "Exploration in sequential recommender systems via graph representations"
2. 12th International Conference on Network Analysis (NET) 2022. 25 May 2022. Topic: "Predicting Molecule Toxicity via Graph Neural Networks"
3. Journal Club of the Artificial Intelligence Research Institute (AIRI), Moscow, Russia, 27 October 2021. Topic: "Temporal graph embeddings".
4. Internal seminar of the Sber Recommender Systems Platform, 26 October 2021. Topic: "Review of the exploration in recommender systems"

5. Internal seminar of the Sber Recommender Systems Platform, 8 July 2021. Topic: "Graph-based recommender systems"
6. Artificial Intelligence Journey, 20 December 2020. Topic: "Eternal Student: How to check whether adaptive recommender system learns"
7. 2nd International Workshop "New Level of Visualization - New Level of Analytics"(NLVNLA-2020), 24 January 2020. Topic: "Visualizing Structural Data via Network Embeddings"
8. International Conference on Analysis of Images, Social Networks and Texts (AIST-2019), 19 July 2019. Topic: "Prediction of new itinerary markets for airlines via network embedding"
9. 9th International Conference on Network Analysis (NET) 2019. 18 May 2022. Topic: "Prediction of new itinerary markets for airlines via network embedding"

## 5.4 Апробация исследований

Модель, предложенная в [20] была протестирована Лабораторией Искусственного Интеллекта Сбера для решения банковских задач. Разработанная модель была применена к графу транзакций между компаниями, чтобы рассчитать их динамические представления. Полученные векторы были использованы в качестве признаков компаний для обучения модели классификации LightGBM [11]. Целевой переменной классификатора являлся факт дефолта компании спустя полгода после последней транзакции в тренировочных данных для обучения представлений. Предложенный подход к извлечению признаков компаний показал сравнимое качество с существующими моделями банка на основе последовательностей и показал значительное улучшение результатов для графовых методов на подвыборке из банковского графа транзакций.

**Личный вклад в статьи.** В статье [17] научный руководитель Илья Макаров и Ловро Шубель поставили задачи, вычитывали текст, проектировали эксперименты и описали основную часть методов; автор диссертации провел эксперименты, сравнил модели, сделал обзор применений и части методов, вывел открытые проблемы и заключения; Никита Никитинский подготовил обзор части применений к задачам обработки естественного языка и вычитывал текст статьи. В [19] научный руководитель Илья Макаров поставил задачи, вычитывал текст и проектировал эксперименты; автор диссертации проводил эксперименты со сложными методами объединения текстовой и структурной информации такими, как GCN, GraphSAGE, GAT и GIC, предложил

методологию для практического использования; Михаил Макаров проводил эксперименты по остальным структурным методам. В работе [18] Илья Макаров поставил проблему, вычитывал текст и проектировал эксперименты; Ксения Коровина разработала основную часть системы; автор диссертации пересмотрел их и улучшил для задачи частичного обучения с учителем модели JONNEE, провел эксперименты и сравнил модели. В статье [20] Илья Макаров, Андрей Савченко и Дмитрий Бабаев поставили проблему, вычитывали текст и проектировали эксперименты; Арсений Коровко и Леонид Шерстюк реализовали основную часть среды и реализовали новые элементы модели, Никита Северин и автор диссертации переписали эту среду до состояния готового продукта, реализовали ведущие методы в рамках нее и провели больше экспериментов, автор диссертации провел финальные эксперименты, выбрал лучшее сочетание компонентов моделей и доработал их, чтобы получить ведущие результаты, вычитывал статью; Александр Михеев проводил эксперименты на внутренних данных Сбера. Статья [16] полностью реализована автором диссертации с небольшой помощью при вычитке текста, консультациями и обсуждениями экспериментов и моделей с Ильей Макаровым.

Автор диссертации является главным (corresponding) автором в статьях [17, 19, 18, 16].

## 6 Содержание работы

Данная секция описывает общие идеи, результаты и новизну отдельных глав (статей) диссертации. Имя подсекции соответствует имени главы (статьи) в основном тексте диссертации.

**Объем и структура работы.** Диссертация состоит из вступления, заключения и содержания пяти статей в приложении. Полный объем работы с приложением – 160 страниц.

### 6.1 Survey on graph embeddings and their applications to machine learning problems on graphs

Первая глава диссертации концентрируется на задаче учета структуры графа при построении непрерывных представлений. Работа рассматривает разные типы графов, в частности графы взаимодействий пользователей и объектов, их применение к задаче рекомендательных систем. Основная цель главы – проанализировать и сравнить существующие модели, используемые для построения непрерывных представлений графов, оценить их эффективность в последующих (downstream) задачах, в частности предсказания связей.

Существует много разных методов обучения непрерывных представлений графов. Классический подход применяет разложения матриц к разным матричным представлениям графа [4, 25, 1, 46, 40, 12]. В этом случае задача построения векторов вершин графа равносильна решению задачи рекомендаций через матричные разложения. Более современные методы предлагают использовать методы построения распределенных представления слов [23] на случайных блужданиях на графе. Такой подход работает из-за похожих степенных законов распределения степеней вершин и встречаемости слов в тексте [29]. Основные отличия методов, использующих случайные блуждания, лежат в построении траекторий обхода графа для сохранения разных свойства [29, 7, 36, 30]. Наиболее современными и популярными подходами являются графовые нейронные сети. Графовые сверточные сети представляют вершины в виде набора нелинейных преобразований признаков и их агрегации по соседям [14]. Такой подход приближает спектральное разложение графа, сохраняя важные структурные свойства. Сети внимания на графах (Graph Attention Networks) используют похожую идею, но применяют механизм внимания, который позволяет выучить более успешные агрегации соседей [43]. Алгоритм передачи сообщений (message-passing framework) обобщает графовые нейронные сети, разбивая процесс обучения векторов вершин на две части: кодирование сообщений и агрегация соседей [6]. Шаг кодирования сообщений под-



разумевают, что алгоритм применяет трансформацию к признакам вершин и кладет их на прилегающее ребро. Шаг агрегации заключается в объединении всех сообщений с ребер в один вектор вершины (ее конечное представление). Такой алгоритм позволяет использовать пакетные (batch) подходы к обучению и предсказанию графовых нейронных сетей. Основная идея GraphSAGE заключается в том, чтобы использовать семплированные окрестности вершины для более эффективного масштабирования нейронных сетей на большие графы. Для генерации графов были адаптированы подходы из других областей: вариационные автокодировщики, рекуррентные модели и генеративно-состязательные сети. Кроме того, подобные подходы вместе с самообучением позволяют строить более устойчивые модели, имеют лучшую обобщающую способность.

Альтернативной точкой зрения на построения непрерывных представлений является взгляд со стороны различных типов графов. Для работы с атрибутивными графами мы можем условно разделить подходы на несколько групп: добавление атрибутов вершин в задачу снижения размерности, построение графа близости по контенту и добавление его структуры к изначальной информации, графовые нейронные сети. Для гетерогенных сетей наиболее классическим подходом является разделение ребер и вершин разных типов на подгруппы и обучение отдельных представлений для каждой из них. Более современные методы на основе случайных блужданий в основном переформулируют задачу оптимизации так, чтобы учитывать близость отдельно по каждому типу связи. Графовые нейронные сети, обычно, используют механизм внимания, для того, чтобы найти оптимальное объединение информации разной природы – близость внутри одной группы вершин и ребер, близость между разными типами, признаки вершин и ребер. Динамические графы рассматриваются с двух позиций: последовательность графов или последовательность событий с вершинами и ребрами. В первом случае строится представление в каждом слепке и сохраняется динамика с помощью рекуррентных моделей. Во втором варианте каждая отдельная вершина, обычно, имеет свою память и обновление применяется конкретно к ней, а сохранение структуры с учетом времени ребра уже используется в самом конце. В больших графах используются два основных варианта: семплирование подграфов, чтобы была возможность итерироваться по графу покомпонентно, и искажение (coarsening) графа, чтобы его было проще разделить на подгруппы и распараллелить расчеты.

В плане задач графового машинного обучения основные отличия между методами кроются в дополнительных целях при обучении. Например, для классификации вершин часто добавляют метрические функции потерь, которые делают вектора вершин одного типа ближе. В задаче кластеризации

часто добавляют штрафы на модулярность после кластеризации векторов с помощью k-means или dbSCAN. В случае задач на уровне графов или подграфов дополнительно применяется агрегация векторов вершин и ребер.

В результате анализа мы выделяем несколько открытых проблем. Во-первых, на момент написания статьи область динамических графов была слабо исследована. Во-вторых, большинство моделей концентрировалось только на представлении вершин и рассматривала ребра только как их агрегацию, поэтому практически отсутствовали работы по учету признаков ребер. Большинство моделей и методов страдает от слабой масштабируемости, а те методы, что могут работать на больших графах склонны к различным смещениям и искажениям структуры, которые слабо изучены. Наконец, наиболее важной проблемой является отсутствие общих моделей и стратегий для выбора оптимального подхода для конкретного графа с конкретными свойствами.

Последняя часть главы посвящена экспериментам для изучения свойств разных типов представлений. Эксперименты нацелены на понимание того, как конкретные свойства графов сопряжены с качеством ведущих моделей. Также модели протестированы на различных генеративных моделях графов, чтобы оценить их свойства в контролируемой среде.

Новизна работы заключается в предложенной методологии анализа и сравнения непрерывных представлений графа для конкретных задач машинного обучения, связанных с графами. Мы обнаружили, что классические структурные непрерывные представления графа с правильным подбором целевой функции под конкретный тип сети позволяют достичь сравнимых результатов с более сложными моделями глубокого обучения в последующих задачах.

В результате данной главы были изучены и сравнены разные подходы к учету структуры графа. Данный шаг является первым необходимым для того, чтобы построить рекомендательную систему на основе графа, которая может учитывать коллаборативные признаки.

Статья опубликована в журнале PeerJ Computer Science. Он индексируется как Q1 (2020) и Q2 (2021) в Scopus.

## **6.2 Fusion of text and graph information for machine learning problems on networks**

Данная глава решает вторую поставленную нами задачу: объединение контентной и коллаборативной информации для решения задачи классификации вершин и предсказания связей (рекомендаций).

Предыдущие исследования непрерывных представлений графов фокусировались только на структурной части проблемы и целились улучшить сохранение именно структурной информации. Поэтому статьи в этой области

использовали только простые стратегии кодирования текстом по типу мешка слов [9] или Tf-Idf [37]. Однако, гораздо более важной проблемой поднятой в предыдущей главе является то, как эффективно балансировать между текстовыми и структурными признаками вершин. Цель этой главы – предложить такие стратегии, которые могут эффективно учитывать текстовые атрибуты вершин в последующих задачах. Таким образом, мы изучаем современные продвинутое методы кодирования текстов: LDA [3], word2vec [24], sent2vec [27], Sentence BERTS [33] и ERNIE [41]. В первую очередь мы пытаемся оценить достаточно ли только текстовой информации для решения задач классификации вершин и предсказания связей. Далее мы анализируем то, как такая информация может быть интегрирована в рамках подхода построения непрерывных представлений графов. Были представлены разные стратегии: наивная смесь (конкатенация) структурных [25, 29, 7] и текстовых представлений, сложные техники разложения матриц [49, 28] и графовые нейронные сети [13, 8, 44, 22]. Более того, графовые нейронные сети также тестировались и валидировались в формате непрерывного решения конкретной задачи (end-to-end), чтобы показать более высокое качество в решении последующих задач.

Мы пришли к следующим выводам. Во-первых, мы обнаружили, что комплексные стратегии кодирования текстов позволяют заметно повысить качество структурных моделей в графовых задачах. Во-вторых, мы показали, что кодирование текстов моделью SBERT сильнее улучшает задачу предсказания связей, а Sent2Vec напротив – классификации вершин. В итоге, мы предложили собственную модификацию архитектуры GCN.

Новизна исследования заключается в новом взгляде на задачи графового машинного обучения и учете современных продвинутых методов кодировки текстов для улучшения качества прикладных задач.

В результате данной главы было показано, что для успешного решения задачи предсказания связей (рекомендаций) недостаточно использовать только структуру графа, нужно также учитывать и признаки. Более того, были проанализированы разные варианты учета такой информации, показана эффективность использования современных техник кодирования информации для графов, аннотированных текстами. Таким графы являются одними из наиболее популярных в современных рекомендательных системах и встречаются в социальных сетях, медиа, описаниях товаров в магазинах и т.д.

Статья опубликована в журнале PeerJ Computer Science. Данный журнал проиндексирован как Q1 (2020) и Q2 (2021) в Scopus.

### 6.3 JONNEE: Joint network nodes and edges embedding

Данная глава сфокусирована на проблеме использования атрибутов ребер и построение эффективных представлений ребер для последующих задач. Большинство работ рассматривает только атрибуты вершин или простые параметры ребер по типу веса. Цель статьи – построить подход, в рамках которого такие атрибуты могут быть эффективно использованы.

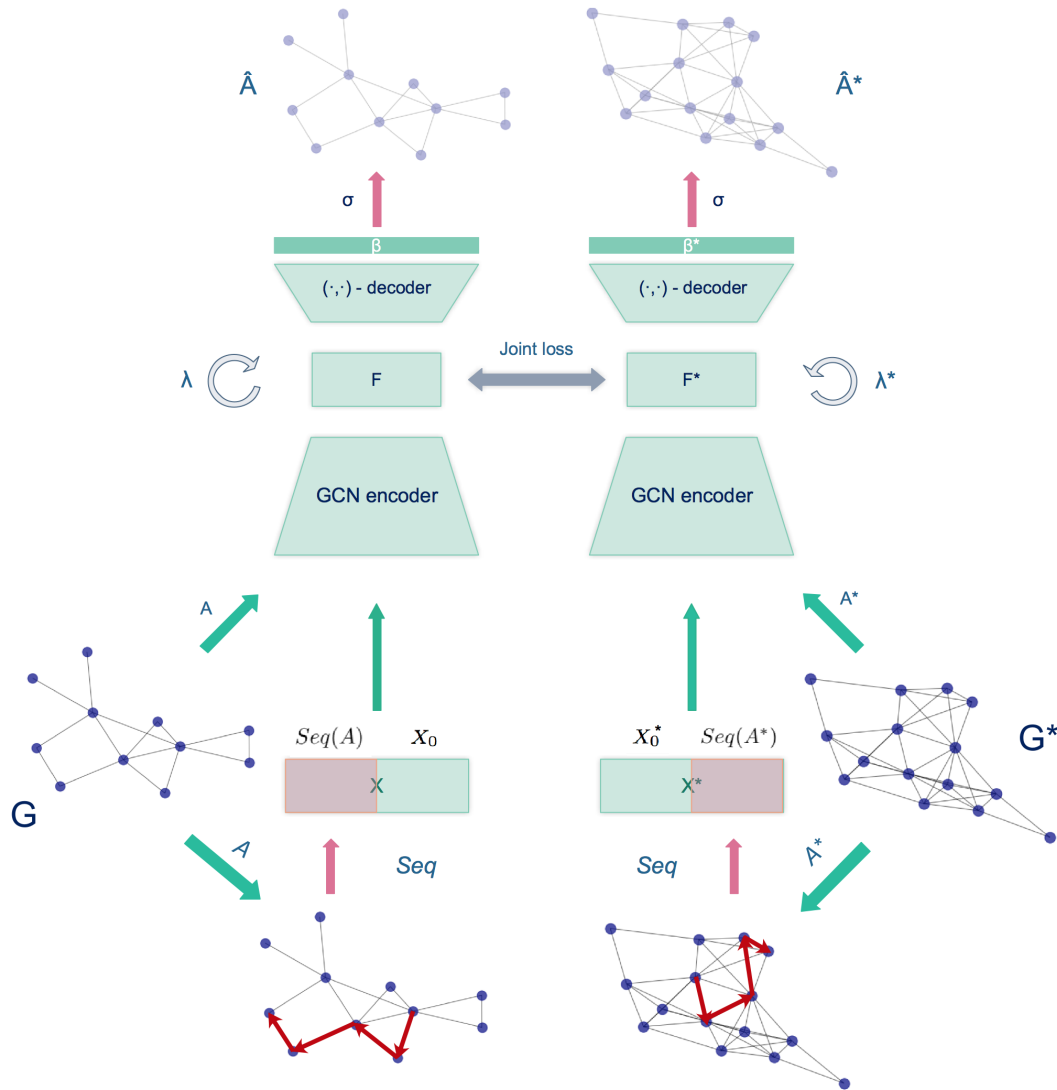


Рис. 6: Архитектура модели JONNEE

Основная идея разработанной модели (рисунок 6 – использовать уже хорошо изученные подходы к обучению представлений вершин в применении

к ребрам. Для того, чтобы этого достичь мы применяем VGAE не только к обычному графу, но и к его двойственному представлению. В двойственной графе роль вершин исполняют ребра, а связи между ними создаются на основе инцидентности с другими ребрами в изначальном графе. Таким образом, используя другое представление графа мы способны применять уже существующие методы, чтобы учитывать информацию о признаках ребер и, в целом, получать более качественные представления. Тем не менее, предложенные кодировщики должны быть согласованы в том, как они представляют разные вершины и ребра. Поэтому мы предложили новый подход к обучению таких моделей на основе самообучения. Мы предлагаем функцию потерь, которая состоит из следующих частей:

1. Реконструкция автокодировщика для двух графов

$$L_G = \|A - \widehat{A}\|_F^2 \sim \frac{1}{|V|^2} \sum_{i,j} (\hat{a}_{ij} - a_{ij})^2 \quad (1)$$

$$L_{G^*} = \|A^* - \widehat{A}^*\|_F^2 \sim \frac{1}{|E|^2} \sum_{i,j} (\hat{a}_{ij}^* - a_{ij}^*)^2 \quad (2)$$

2. Совместная функция потерь ( $f$  и  $f^*$  – непрерывные представления для вершин в обычном и двойственном графах соответственно,  $N(v)$  – набор соседей вершины)

$$L_{G^* \rightarrow G} = \sum_{v \in V} \left\| f(v) - \frac{1}{|N^G(v)|} \sum_{u \in N(v)} f^*((u, v)) \right\|_2^2 \quad (3)$$

$$L_{G \rightarrow G^*} = \sum_{e=(u,v) \in E} \left\| f^*(e) - \frac{\sum_{t \in N^G(u) \cup N^G(v)} f(t)}{|N^{G^*}(e)|} \right\|_2^2 \quad (4)$$

3.  $L_2$ -регуляризации для двух графов

Наиболее важная часть – формулы 3 и 4. Их основная идея заключается в том, что мы хотим, чтобы представление двух кодировщиков были близки между собой для одинаковых сущностей. Мы стягиваем представление ребер в дуальном графе, соответствующих вершине в обычном, к ее значению в изначальном графе и наоборот. Формулы 1 и 2 отвечают за то, чтобы сами по себе кодировщики успешно справлялись с задачей реконструкции графа. Последний пункт функции потерь штрафует модель за излишнюю сложность, чтобы снизить потенциал к переобучению.

Не смотря на то, что полученная архитектура позволяет обучать более качественные представления связей, она является также и более сложной вычислительно, так как появляется второй кодировщик, а двойственный граф обычно значительно больше обычного. Чтобы снизить вычислительную сложность модели и ускорить ее сходимость мы иницилируем векторами, полученными с помощью подходов на основе блужданий: `node2vec` или `diff2vec`.

Предложенный подход показывает сравнимое качество с другими ведущими моделями глубокого обучения при частичном обучении с учителем в задачах предсказания связей и классификации вершин. Модель имеет высокое качество как в случае обучения без учителя, так и в случае частичного обучения с учителем. Она дает четкую кластерную структуру при визуализации графов.

Новизна модели JONNEE заключается в двух факторах. Во-первых, она улучшает классические вариационные автокодировщики с помощью стратегий последовательной кодировки вершин. Во-вторых, она тренируется с помощью новой стратегии самообучения, которая притягивает представления вершин и их двойственных копию в дуальном графе.

Модель JONNEE является важным шагом к построению более качественных рекомендательных систем. Во-первых, задача предсказания связей зависит от того, чтобы на вход были поданы качественные представления ребра. Во-вторых, данная модель дает возможность учитывать контекст действия пользователя, который часто является важным в практических применениях, например, обычно, люди слушают разную музыку, когда работают и на вечеринках, люди смотрят разные фильмы по одиночке в дороге и вечером вместе с детьми.

Статья опубликована в IEEE Access. Этот журнал проиндексирован как Q1 в Scopus и Web of Science.

## 6.4 Temporal network embedding framework with causal anonymous walks representations

Данная глава посвящена проблеме учета динамики графов. Как уже говорилось выше, пользователь имеет свойство менять поведение, имеет некоторую причинность в своих действиях, поэтому проблема темпоральности является одной из наиболее важных в рекомендациях. Основная цель главы – изучить методы учета динамики графа, проанализировать их качество и предложить улучшения для более качественной работы на динамических графах, в частности графов взаимодействий пользователей и объектов.

В этой главе мы разбираем новые непрерывные представления графов, основанные на временных графовых сетях (Temporal Graph Network, TGN)

[35] и анонимизированных причинных блужданиях (Causal Anonymous Walks, CAW) [47]. Мы выбрали TGN в качестве основы новой модели, потому что она обобщает большинство существующих подходов непрерывных представлений динамических графов и имеет гибкую модульную архитектуру. Она позволяет быстро и выразительно обновлять память вершины. CAW предлагает противоположный взгляд на проблему эволюции графа. Авторы этой модели отказываются от идеи памяти, поэтому анонимизируют каждую вершину. Вместо памяти, они предлагают такую модель, которая может неявно использовать законы эволюции конкретного графа, игнорируя индексы вершин. Из-за этого, CAW не может делать представления отдельных вершин, но зато позволяет значимо лучше учитывать структуру и эволюционные мотивы графа. Таким образом, мы предлагаем объединить два идейно противоположных подхода, чтобы покрыть проблемы каждого и, в результате, построить более точный метод кодирования динамических графов. Для того, чтобы учитывать признаки CAW, мы конкатенируем их к представлением ребра TGN и обновляем память с их учетом.

Второй важный вклад данной работы – стандартизированный фреймворк для обучения и валидации непрерывных представлений динамических графов. Он предоставляет стандартизированные обертки для подготовки графов, выбора типов и параметров пакетов, разделения данных на трансдуктивные и индуктивные (решение проблемы холодного старта) подвыборки, стандартные гибкие интерфейсы для построения моделей и единый процесс тренировки и оценки качества моделей.

Новизна работы состоит из двух пунктов. Во-первых, мы предлагаем новую модель, которая решает проблемы предыдущих TGN и CAW, объединяя их архитектуры. Во-вторых, мы предлагаем единую среду для оценки качества моделей векторизации динамических графов для решения последующих задачах. Он позволяет гибко интегрировать разные модели и динамические графы в рамках единого подхода.

Дополнительно, мы показываем эффективность предложенной модели и ее подмодулей через поэлементное удаление и тестирование всего процесса. Также, мы предлагаем промышленное применение подхода к задаче предсказания банкротства на транзакционном графе крупного европейского банка. Мы показали, что расширение признаков динамического внимания поверх временных случайных блужданий улучшает количественные и качественные результаты реальных применений на банковских данных.

Эксперименты показали возможность использования метода для решения задач предсказания связей или классификации вершин динамических графов и значимо улучшить качество работы на таких задачах.

Учет динамики графа позволяет строить рекомендательные системы, кото-

рые улавливают причинность и динамику в поведении пользователя, способны адаптироваться к изменениям в его поведении.

Статья опубликована в журнале PeerJ Computer Science. Данный журнал проиндексирован как Q1 (2020) и Q2 (2021) в Scopus.

## 6.5 Exploration in sequential recommender systems via graph representations

Последняя работа посвящена исследованию неизвестных состояний (пар пользователь-объект) и поиску баланса между использованием выученных зависимостей и поиском потенциально новых изменений. Данная проблема заключается в том, что из-за изменений поведения пользователя текущая структура графа становится менее релевантной, поэтому качество непрерывных представлений и рекомендаций ухудшается. Цель данной главы разработать новые стратегии исследования неизвестных состояний на основе графового представления задачи рекомендаций. В ней мы адаптировали подход внутренней мотивации из обучения с подкреплением [50] и применили его к графовым данным.

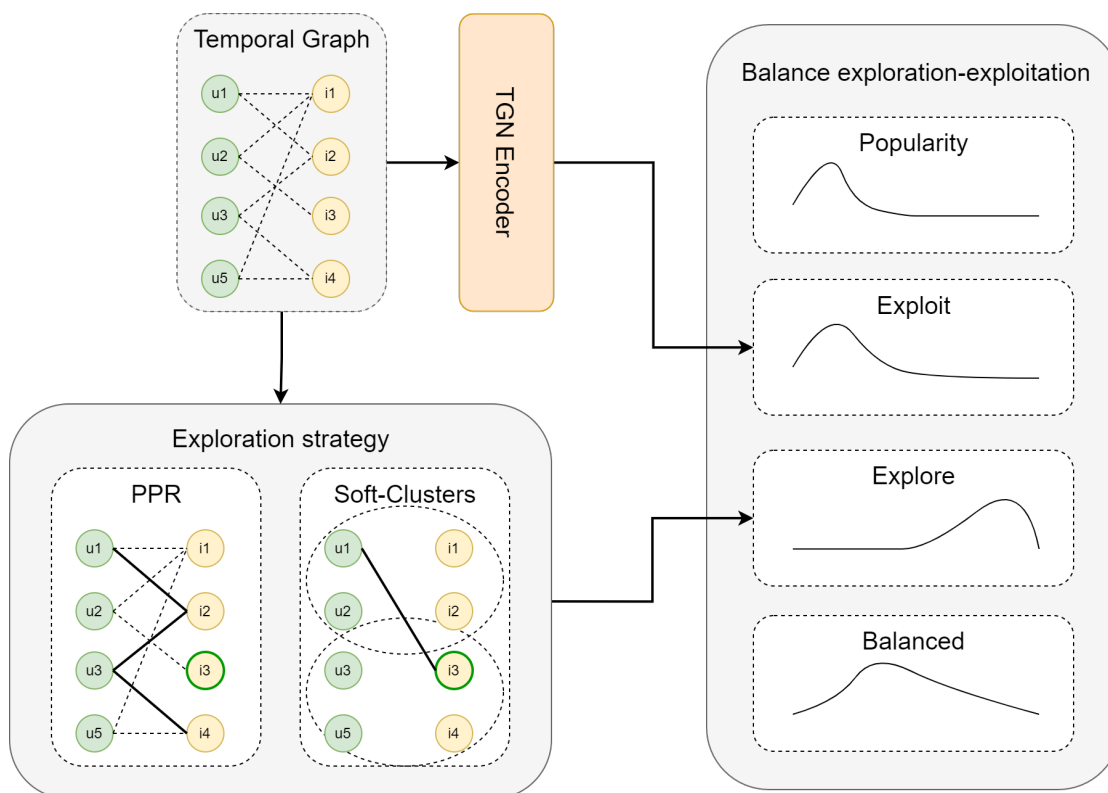


Рис. 7: Методы исследования неизвестных пар пользователь-объект



Важным свойством графа взаимодействий является то, что он подчиняется степенному закону (power law). Таким образом, связи внутри графа тяготеют к наиболее популярным вершинам (рисунок 7). В результате, когда мы учим модели рекомендаций, предсказания смещены в сторону более популярных объектов. Глобально, в этом нет ничего страшного, тем не менее, из-за изменения в поведении пользователей или трендах, старые наиболее популярные вершины становятся неактуальными. Исходя из этого мы предлагаем стратегию на основе персонализированного пейджранка (PPR) [26], которая оценивает локальную популярность объектов в ближайшей окрестности пользователя. Мы модифицировали PPR, чтобы он учитывал только относительно недавнюю временную динамику. Для этого, мы семплируем короткие (длина 3) каузальные (каждое следующее ребро возникло раньше предыдущего) случайные блуждания. Далее, мы считаем встречаемость каждого объекта и добавляем вес обратно пропорциональный ей как бонус к предсказаниям TGN.

Второй метод исследования основан на идее маленького мира. Несмотря на то, что изначально графы взаимодействий являются достаточно плотными, мы сталкиваемся с проблемой, что недавно добавленные объекты и пользователи являются труднодоступными на этом графе. Таким образом, при агрегации окрестности вершин учесть вероятность учесть новые крайне мала. Поэтому мы хотели бы строить такие связи в графе, которые соединяют далекие вершины и была возможность быстрее находить более релевантные товары для пользователя. Для того, чтобы добиться такого свойства мы выделяем сообщества на графе и соединяем вершины из относительно далеких групп. Кластеризация графов является вычислительно сложной задачей, поэтому мы упрощаем ее до поиска "мягких" кластеров. Суть мягкой кластеризации заключается в том, что мы с помощью GNN проецируем каждую вершину в пространство фиксированной размерности (128), нормализуем его через софтмакс и используем полученный вектор, как распределения по кластерам. Далее мы добавляем к предсказаниям TGN бонус пропорциональный вероятности принадлежности к кластеру для пользователя и обратно пропорциональный для объекта.

Результаты экспериментов показывают важность исследования в задачах онлайн-адаптации моделей. Если в данных представлена правильная временная структура, разные подходы к исследованию улучшают качество моделей. Тем не менее, качество стратегий исследования зависит от свойств данных. Для графов, где представлено относительно мало положительных ребер лучше использовать подход на основе персонализированного пейджранка. В такой ситуации он семплирует меньше уникальных вершин, что позволяет оценить локальную популярность куда лучше. В противоположной ситуации, когда

средняя степень вершины высокая, а граф плотный, метод исследования, основанный на мягкой кластеризации графа (маленький мир) показывает качество лучше.

Предложенные стратегии показывают сравнимое качество с другими методами исследования для рекомендательных систем. В будущем, мы целимся применить предложенные подходы к задачам повторного потребления, мы также планируем изучить предложенные подходы в более сложных средах, например, многокомпонентные рекомендательные системы. Также, важно оценить качество таких методов в случае гетерогенных связей и вершин, например, когда разных типов могут быть пользователи (с платной подпиской или без) или объекты (фильмы, сериалы, блоги) и связи между ними (покупка, просмотр, лайк).

Статья находится на рецензии в журнал IEEE Access, Q1.

## 7 Заключение

Диссертация основана на опубликованных работах [17, 19, 18, 20] и статье [16], которая находится на рецензии в журнал IEEE Access. Статьи [17, 19] анализируют текущие достижения в области и проводят эмпирические исследования ведущих моделей непрерывных представлений графов и их объединения с текстовыми моделями для того, чтобы вывести ряд советов для практического применения моделей. Статьи [18, 20, 16] предлагают новые модели и методы для решения важных проблем рекомендательных систем, основанных на графах: учет характеристик контекста (ребра), учет последовательности действий, петли обратной связи и сдвиги в распределениях данных. В совокупности все статьи помогают решить основные проблемы рекомендательных систем в рамках одного подхода графового машинного обучения.

Основной вклад диссертации, выносимый на защиту:

1. Мы предложили детальную таксономию подходов к построению непрерывных представлений графов и их применений [17]
2. Мы провели оценку ведущих подходов к построению непрерывных представлений графов для задачи предсказания связей [17]
3. Мы составили методологию автоматического извлечения свойств графов с текстовыми атрибутами вершин для задачи предсказания связей [19]
4. Мы создали новый подход к совместному построению представлений вершин и ребер, основанный на самообучении с использованием соответствующего дуального представления графа [18]
5. Мы предложили новый подход к построению темпоральных непрерывных представлений графа показывающий ведущие результаты в различных темпоральных задачах графового машинного обучения [20]
6. Мы реализовали стандартизированную среду для обучения и валидации темпоральных непрерывных представлений графа, сравнили ведущие модели в рамках этой среды, и получили новое понимание реального качества моделей в сравнении с оригинальными результатами [20]
7. Мы спроектировали новую стратегию исследования неизвестных состояний на основе Personalized PageRank [26] [16]
8. Мы предложили новую стратегию исследования неизвестных состояний на основе идеи маленького мира [26] [16]
9. Мы составили методологию применения предложенных методов к задаче рекомендательных систем с онлайн адаптацией в динамичных средах [16]

**Направления будущих исследований.** Будущие исследовательские проблемы могут быть выведены из результатов описанных работ. Одним из важных выводов статьи [17] является недостаток мета-стратегий для подбора актуальных методов векторизации графов для конкретных задач в AutoML стиле. Статья [19] может быть расширена применением техник самообучения и функций потерь для более качественного объединения текстовой и структурной информации. Другой многообещающий путь – переформулировать данную задачу в терминах поиска информации (например, быстрые модели первого уровня в многокомпонентных рекомендательных системах) или в терминах обработки естественного языка (например, [51]). Основным недостатком модели JONNEE [18] является слабая масштабируемость. Для улучшения работы с большими графами можно изменить основные блоки модели на более быстрые подходы на основе подвыборок. Другое направление развития – проанализировать качество JONNEE на классических данных рекомендательных систем с контекстной информацией. Статья [20] может быть улучшена также благодаря применению техник масштабируемости. Стратегии исследования неизвестных состояний [16] подвержены проблеме излишнего сглаживания (over-smoothing). То есть они могут быть улучшены применением дифференцируемых групповых нормализаций [52] или разделением графа на разные подграфы с учетом текущего контекста пользователя. Другим направлением для [16] является работа с гетерогенными ребрами и вершинами для разных типов взаимодействий между пользователями и объектами, например, негативные и позитивные связи, или покупки против добавления в корзину.

## Список литературы

- [1] Sami Abu-El-Haija, Bryan Perozzi, and Rami Al-Rfou. Learning edge representations via low-rank asymmetric projections. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM '17*, page 1787–1796, New York, NY, USA, Nov 2017. Association for Computing Machinery.
- [2] Gediminas Adomavicius, Jesse C Bockstedt, Shawn P Curley, and Jingjing Zhang. Effects of online recommendations on consumers' willingness to pay. *Information Systems Research*, 29(1):84–102, 2018.
- [3] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3:993–1022, 2003.
- [4] Shaosheng Cao, Wei Lu, and Qionghai Xu. Grarep: Learning graph representations with global structural information. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, CIKM '15*, page 891–900, New York, NY, USA, 2015. Association for Computing Machinery.
- [5] João Gama, Indrė Žliobaitė, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. A survey on concept drift adaptation. *ACM computing surveys (CSUR)*, 46(4):1–37, 2014.
- [6] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Message passing neural networks. In *Machine learning meets quantum physics*, pages 199–214. Springer, 2020.
- [7] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. *arXiv preprint arXiv:1607.00653*, 1607.00653, 2016.
- [8] William L Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Proceedings of IC on NIPS'17, NIPS'17*, pages 1025–1035. Curran Associates Inc., 2017.
- [9] Zellig Harris. Distributional structure. *Word*, 10(2-3):146–162, 1954.
- [10] Dietmar Jannach and Michael Jugovac. Measuring the business value of recommender systems. *ACM Trans. Manage. Inf. Syst.*, 10(4), dec 2019.
- [11] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. LightGBM: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30:3146–3154, 2017.

- [12] Zekarias T. Kefato, Nasrullah Sheikh, and Alberto Montresor. Which way? direction-aware attributed graph embedding, 2020.
- [13] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 1609.02907, 2016.
- [14] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 1609.02907, 2017.
- [15] Dmitrii Kiselev and Ilya Makarov. Prediction of new itinerary markets for airlines via network embedding. In *International Conference on Analysis of Images, Social Networks and Texts*, pages 315–325. Springer, 2019.
- [16] Dmitrii Kiselev and Ilya Makarov. Exploration in sequential recommender systems via graph representations (on review). *IEEE Access*, 2022.
- [17] Ilya Makarov, Dmitrii Kiselev, Nikita Nikitinsky, and Lovro Subelj. Survey on graph embeddings and their applications to machine learning problems on graphs. *PeerJ Computer Science*, 7:e357, 2021.
- [18] Ilya Makarov, Ksenia Korovina, and Dmitrii Kiselev. JONNEE: Joint network nodes and edges embedding. *IEEE Access*, 2021.
- [19] Ilya Makarov, Mikhail Makarov, and Dmitrii Kiselev. Fusion of text and graph information for machine learning problems on networks. *PeerJ Computer Science*, 7, 2021.
- [20] Ilya Makarov, Andrey Savchenko, Arseny Korovko, Leonid Sherstyuk, Nikita Severin, Dmitrii Kiselev, Aleksandr Mikheev, and Dmitrii Babaev. Temporal network embedding framework with causal anonymous walks representations. *PeerJ Computer Science*, 8:e858, 2022.
- [21] Masoud Mansoury, Himan Abdollahpouri, Mykola Pechenizkiy, Bamshad Mobasher, and Robin Burke. Feedback loop and bias amplification in recommender systems. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 2145–2148, 2020.
- [22] Costas Mavromatis and George Karypis. Graph infoclust: Leveraging cluster-level node information for unsupervised graph representation learning, 2020.
- [23] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013.

- [24] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 1301.3781, 2013.
- [25] Mingdong Ou, Peng Cui, Jian Pei, Ziwei Zhang, and Wenwu Zhu. Asymmetric transitivity preserving graph embedding. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 1105–1114, New York, NY, USA, Aug 2016. Association for Computing Machinery.
- [26] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.
- [27] Matteo Pagliardini, Prakhar Gupta, and Martin Jaggi. Unsupervised learning of sentence embeddings using compositional n-gram features. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 528–540. Association for Computational Linguistics, 2017.
- [28] Shirui Pan, Jia Wu, Xingquan Zhu, Chengqi Zhang, and Yang Wang. Tri-party deep network representation. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, page 1895–1901. AAAI Press, 2016.
- [29] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of ACM SIGKDD IC on KDD'20*, pages 701–710. ACM, 2014.
- [30] Bryan Perozzi, Vivek Kulkarni, and Steven Skiena. Walklets: Multiscale graph embeddings for interpretable network classification, 2016.
- [31] Idris Rabi'u, Naomie Salim, Aminu Da'u, and Akram Osman. Recommender system based on temporal models: A systematic review. *Applied Sciences*, 10(7), 2020.
- [32] Shaina Raza and Chen Ding. Progress in context-aware recommender systems—an overview. *Computer Science Review*, 31:84–97, 2019.
- [33] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*,

pages 3982–3992, Hong Kong, China, November 2019. Association for Computational Linguistics.

- [34] Steffen Rendle. Factorization machines. In *2010 IEEE International conference on data mining*, pages 995–1000. IEEE, 2010.
- [35] Emanuele Rossi, Ben Chamberlain, Fabrizio Frasca, Davide Eynard, Federico Monti, and Michael Bronstein. Temporal graph networks for deep learning on dynamic graphs. *arXiv preprint arXiv:2006.10637*, 2006.10637, 2020.
- [36] Benedek Rozemberczki and Rik Sarkar. Fast sequence-based embedding with diffusion graphs. In *International Workshop on Complex Networks*, pages 99–107. Springer International Publishing, 2018.
- [37] Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5):513–523, 1988.
- [38] Rachna Sethi and Monica Mehrotra. Cold start in recommender systems—a survey from domain perspective. In Jude Hemanth, Robert Bestak, and Joy Jong-Zong Chen, editors, *Intelligent Data Communication Technologies and Internet of Things*, pages 223–232, Singapore, 2021. Springer Singapore.
- [39] Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer. In *Proceedings of the 28th ACM international conference on information and knowledge management*, pages 1441–1450, 2019.
- [40] Jiankai Sun, Bortik Bandyopadhyay, Armin Bashizade, Jiongqian Liang, P Sadayappan, and Srinivasan Parthasarathy. Atp: Directed graph embedding with asymmetric transitivity preservation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 265–272. AAAI, 2019.
- [41] Yu Sun, Shuohuan Wang, Yukun Li, Shikun Feng, Hao Tian, Hua Wu, and Haifeng Wang. Ernie 2.0: A continual pre-training framework for language understanding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8968–8975, 2020.
- [42] Gábor Takács and Domonkos Tikk. Alternating least squares for personalized ranking. In *Proceedings of the sixth ACM conference on Recommender systems*, pages 83–90, 2012.



- [43] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 1710.10903, 2017.
- [44] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.
- [45] Shoujin Wang, Liang Hu, Yan Wang, Longbing Cao, Quan Z Sheng, and Mehmet Orgun. Sequential recommender systems: challenges, progress and prospects. *arXiv preprint arXiv:2001.04830*, 2019.
- [46] Xiao Wang, Peng Cui, Jing Wang, Jian Pei, Wenwu Zhu, and Shiqiang Yang. Community preserving network embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.
- [47] Yanbang Wang, Yen-Yu Chang, Yunyu Liu, Jure Leskovec, and Pan Li. Inductive representation learning in temporal networks via causal anonymous walks. *arXiv preprint arXiv:2105.02315*, 2101.05974, 2021.
- [48] Shiwen Wu, Fei Sun, Wentao Zhang, Xu Xie, and Bin Cui. Graph neural networks in recommender systems: a survey. *ACM Computing Surveys (CSUR)*, 2020.
- [49] Cheng Yang, Zhiyuan Liu, Deli Zhao, Maosong Sun, and Edward Y. Chang. Network representation learning with rich text information. In *Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI'15*, page 2111–2117. AAAI Press, 2015.
- [50] Tianpei Yang, Hongyao Tang, Chenjia Bai, Jinyi Liu, Jianye Hao, Zhaopeng Meng, and Peng Liu. Exploration in deep reinforcement learning: A comprehensive survey. *arXiv preprint arXiv:2109.06668*, 2021.
- [51] Michihiro Yasunaga, Jure Leskovec, and Percy Liang. Linkbert: Pretraining language models with document links. *arXiv preprint arXiv:2203.15827*, 2022.
- [52] Kaixiong Zhou, Xiao Huang, Yuening Li, Daochen Zha, Rui Chen, and Xia Hu. Towards deeper graph neural networks with differentiable group normalization. *Advances in Neural Information Processing Systems*, 33:4917–4928, 2020.
- [53] Yu Zhu, Hao Li, Yikang Liao, Beidou Wang, Ziyu Guan, Haifeng Liu, and Deng Cai. What to do next: Modeling user behaviors by time-lstm. In *IJCAI*, volume 17, pages 3602–3608, 2017.