

Национальный исследовательский Нижегородский
государственный университет имени Н. И. Лобачевского

(Институт информационных технологий, математики и механики)

Кафедра: Алгебра, геометрия и дискретная математика (АГДМ)

Обзор статьи:

«Neural Sheaf Diffusion: A Topological Perspective on Heterophily and
Oversmoothing in GNNs»

(Нейронная диффузия на пучках: топологический взгляд на гетерофилию и
oversmoothing в графовых нейронных сетях)

Авторы: Cristian Bodnar, Francesco Di Giovanni, Benjamin P. Chamberlain, Pietro Liò,
Michael M. Bronstein

(arXiv:2202.04579v4 [cs.LG], NeurIPS 2022)

Выполнил:
Студент группы 382351ФИ1
Гусев Дмитрий Алексеевич

Научный руководитель:
Мокеев Дмитрий Борисович

22 ноября 2025 г.

Содержание

Введение	3
Проблема мультимодальных эмбеддингов и её актуальность	3
Графовые нейронные сети и их ограничения	3
Основной результат и вклад статьи	3
Контекст и связанные работы	4
Классические графовые нейронные сети	4
Теория пучков и алгебраическая топология	4
Sheaf Laplacian и диффузия на пучках	5
Ключевые идеи нового подхода	6
Иерархия пучков и выразительная сила	6
Гармоническое пространство и спектральный анализ	6
Обучение пучков из данных	6
Neural Sheaf Diffusion Model	7
Архитектура модели	7
Типы SheafLearner'ов	7
Вычислительная сложность	8
Экспериментальные результаты	8
Синтетические эксперименты	8
Эксперименты на реальных данных	8
Обоснование актуальности для задачи мультимодальных пучков	9
Почему пучки подходят для мультимодальных данных	9
Как Neural Sheaf Diffusion помогает решить задачу	9
Связь с задачей курсовой работы	9
Практические приложения и направления исследований	10
Применение к рекомендательным системам	10
Направления дальнейших исследований	10
Заключение и значимость	10
Математические результаты и доказательства	11
Теоретические результаты о выразительной силе	11
Контроль над асимптотическим поведением	11
Спектральный анализ и неравенство Cheeger	11
Детальное описание экспериментальных результатов	12
Синтетические эксперименты	12
Результаты на реальных датасетах	12
Сравнение с baseline-методами	13
Связанные работы и обсуждение	13
Sheaf Neural Networks и обучение пучков	13
Гетерофилия и Oversmoothing	13
Категориальная теория и GNN	13

План реализации для курсовой работы	13
Этап 1: Реализация базового алгоритма для одной модальности	13
Этап 2: Расширение до мультимодального случая	14
Этап 3: Эксперименты и оценка	14
Заключение и значимость	15

Введение

Проблема мультимодальных эмбеддингов и её актуальность

В современной экосистеме VK существует ряд продуктов (например, VK Видео), где рекомендации формируются на основе контентных эмбеддингов разных модальностей — видео, текста, заголовков и речи. Эмбеддинг — это векторное представление объекта, которое кодирует его семантические свойства. Например, видео о котиках и текст «милые котята» должны иметь похожие эмбеддинги, чтобы система могла понять их связь. Поэтому задача построения репрезентативных мультимодальных эмбеддингов является особенно актуальной.

Традиционные подходы к объединению мультимодальных данных, такие как простое конкатенирование (склеивание векторов) или усреднение эмбеддингов, часто теряют важную структуру данных и семантические связи между различными модальностями. Например, при конкатенировании эмбеддинг видео и эмбеддинг текста просто объединяются в один длинный вектор, но при этом теряется информация о том, как эти модальности должны взаимодействовать. Необходим более совершенный механизм, который учитывает природу различных типов контента и их взаимодействие.

Графовые нейронные сети и их ограничения

Графовые нейронные сети (GNN) стали популярным инструментом для работы с реляционными данными и находят применение в различных областях: от социальных наук и физики частиц до структурной биологии и дизайна лекарств. GNN работают с данными, представленными в виде графа, где вершины соответствуют объектам (например, пользователям или видео), а рёбра — связям между ними (например, просмотрам или лайкам).

Однако в работе с GNN часто наблюдаются две фундаментальные проблемы:

- **Плохая производительность на гетерофильных графах:** Многие GNN построены на сильном предположении гомофилии, т.е. что вершины склонны соединяться с похожими вершинами. Например, в социальной сети друзья обычно имеют схожие интересы. Однако это не всегда верно в реальных данных. В рекомендательных системах видео о котиках может быть связано с видео о собачках не потому, что они похожи, а потому, что пользователи, смотревшие одно, часто смотрят и другое. Такие графы называются гетерофильными, и классические GNN на них работают плохо.
- **Проблема oversmoothing:** При увеличении глубины GNN (количество слоёв) признаки вершин начинают становиться всё более похожими друг на друга, теряя уникальную информацию. Это происходит потому, что каждый слой усредняет признаки соседних вершин, и после многих таких усреднений все признаки становятся практически одинаковыми. В результате модель теряет способность различать разные вершины.

Основной результат и вклад статьи

Статья «Neural Sheaf Diffusion: A Topological Perspective on Heterophily and Oversmoothing in GNNs» (arXiv:2202.04579v4) группы исследователей (Cristian Bodnar, Francesco Di Giovanni, Benjamin P. Chamberlain, Pietro Liò, Michael M. Bronstein) представляет новый класс графовых нейронных сетей, основанный на теории пучков из алгебраической топологии.

Основной результат работы — демонстрация того, что эти две фундаментальные проблемы связаны общей причиной: тривиальной «геометрией» графа, используемой в классических GNN. Решение заключается в использовании нетривиальных пучков на графах, которые позволяют более гибко контролировать процесс передачи информации.

Статья показывает, что:

- Пучковая структура позволяет избежать проблем oversmoothing и гетерофилии
- Дискретизированные параметрические процессы диффузии имеют больший контроль над асимптотическим поведением, чем классические GCN
- Пучки могут быть обучены из данных, что делает подход применимым на практике
- Полученные модели показывают конкурентоспособные результаты как на гетерофильных, так и на гомофильтых графах

Контекст и связанные работы

Классические графовые нейронные сети

Классические GNN, такие как Graph Convolutional Networks (GCN) и Graph Attention Networks (GAT), можно рассматривать как использующие тривиальную структуру пучка: все вершины и рёбра имеют одинаковые векторные пространства, а отображения между ними являются тождественными (то есть просто копируют информацию без преобразования). Это означает, что информация передаётся между вершинами одинаковым образом, независимо от их свойств.

С математической точки зрения, GCN можно интерпретировать как процесс тепловой диффузии. Представьте, что признаки вершин — это температура в разных точках, и тепло распространяется по рёбрам графа, выравнивая температуру. Уравнение тепловой диффузии и его дискретизация методом Эйлера с единичным шагом:

$$\dot{X}(t) = -\Delta_0 X(t) \Rightarrow X(t+1) = X(t) - \Delta_0 X(t) = (I - \Delta_0)X(t)$$

где Δ_0 — это нормализованный граф Лапласиан, который измеряет разницу между признаками соседних вершин.

Сравнивая это с моделью Graph Convolutional Network, мы видим, что GCN является расширенным процессом тепловой диффузии с дополнительной весовой матрицей W (которая позволяет моделировать нелинейные преобразования) и нелинейностью σ (например, ReLU):

$$\text{GCN}(X, A) := \sigma(D^{-1/2} A D^{-1/2} X W) = \sigma((I - \Delta_0)X W)$$

где A — матрица смежности графа, D — диагональная матрица степеней вершин.

С этой точки зрения неудивительно, что GCN особенно подвержен проблемам гетерофилии и oversmoothing: тепловая диффузия по определению стремится выровнять значения, делая признаки соседних вершин всё более похожими. Это хорошо для гомофильтых графов, где похожие вершины должны иметь похожие признаки, но плохо для гетерофильных, где соседние вершины могут быть разными.

Теория пучков и алгебраическая топология

Пучки на графах (cellular sheaves) — это математические объекты из алгебраической топологии, которые позволяют наделить граф более богатой структурой. В отличие от классического подхода, где все вершины и рёбра обрабатываются одинаково, пучок сопоставляет каждому элементу графа своё векторное пространство и задаёт правила преобразования информации при переходе между ними.

Конкретно, пучок сопоставляет:

- Векторное пространство каждой вершине графа

- Векторное пространство каждому ребру графа
- Линейное отображение между пространством вершины и пространством инцидентного ей ребра

Это позволяет, например, иметь разные размерности признаков для разных вершин или преобразовывать информацию при передаче по рёбрам, что даёт гораздо больше гибкости, чем классические GNN.

Клеточный пучок

Определение 1. Клеточный пучок (G, \mathcal{F}) на неориентированном графе $G = (V, E)$ состоит из:

- Векторного пространства $\mathcal{F}(v)$ для каждой вершины $v \in V$
- Векторного пространства $\mathcal{F}(e)$ для каждого ребра $e \in E$
- Линейного отображения $\mathcal{F}_{v \leq e} : \mathcal{F}(v) \rightarrow \mathcal{F}(e)$ для каждой инцидентной пары $v \leq e$

Векторные пространства вершин и рёбер называются *стеблями* (stalks), а линейные отображения — *restriction maps*.

Пространство всех пространств, ассоциированных с вершинами графа, называется пространством 0-коцепей:

$$C^0(G; \mathcal{F}) := \bigoplus_{v \in V} \mathcal{F}(v)$$

Особенно важным подпространством является пространство глобальных сечений:

$$H^0(G; \mathcal{F}) := \{x \in C^0(G; \mathcal{F}) : \mathcal{F}_{v \leq e} x_v = \mathcal{F}_{u \leq e} x_u \text{ для всех рёбер } e = (v, u)\}$$

Это пространство содержит сигналы, которые согласованы с restriction maps пучка вдоль всех рёбер.

Sheaf Laplacian и диффузия на пучках

Для пучка (G, \mathcal{F}) можно определить оператор sheaf Laplacian, который измеряет «несогласованность мнений» в каждой вершине:

Sheaf Laplacian

Определение 2. Sheaf Laplacian пучка (G, \mathcal{F}) — это линейное отображение $L_{\mathcal{F}} : C^0(G, \mathcal{F}) \rightarrow C^0(G, \mathcal{F})$, определённое покоординатно как:

$$L_{\mathcal{F}}(x)_v := \sum_{v, u \leq e} \mathcal{F}_{v \leq e}^\top (\mathcal{F}_{v \leq e} x_v - \mathcal{F}_{u \leq e} x_u)$$

Sheaf Laplacian является положительно полуопределённой блочной матрицей. Нормализованная версия задаётся как $\Delta_{\mathcal{F}} := D^{-1/2} L_{\mathcal{F}} D^{-1/2}$, где D — блочная диагональ $L_{\mathcal{F}}$.

Процесс диффузии на пучках описывается уравнением:

$$X(0) = X, \quad \dot{X}(t) = -\Delta_{\mathcal{F}} X(t)$$

В пределе времени каждая канал признаков проецируется в $\ker(\Delta_{\mathcal{F}})$ — пространство гармонических сигналов, которые согласованы с restriction maps пучка вдоль всех рёбер.

Ключевые идеи нового подхода

Иерархия пучков и выразительная сила

Авторы рассматривают иерархию всё более общих пучков, начиная с тривиального:

- **Тривиальный пучок:** Все стебли изоморфны \mathbb{R} , restriction maps — тождественные. Это соответствует классическому графу Лапласиану.
- **Диагональные пучки:** Restriction maps являются диагональными матрицами. Упрощают вычисления, но ограничивают взаимодействие между измерениями.
- **Ортогональные пучки ($O(d)$ -bundles):** Restriction maps являются ортогональными матрицами. Обеспечивают геометрическую интерпретацию и сохранение структуры данных.
- **Общие пучки:** Restriction maps — произвольные матрицы. Обеспечивают максимальную гибкость, но могут быть склонны к переобучению.

Авторы показывают, что по мере усложнения структуры пучка процесс диффузии на пучках может решать всё более сложные задачи классификации вершин в пределе бесконечного времени. В частности, они доказывают, что при правильной структуре пучка можно избежать проблем oversmoothing и гетерофилии.

Гармоническое пространство и спектральный анализ

Ключевую роль в анализе играет изучение гармонического пространства $\ker(\Delta_{\mathcal{F}})$ с спектральной точки зрения. Авторы предоставляют новое неравенство типа Cheeger для спектрального зазора sheaf Laplacian, что может представлять независимый интерес для спектральной теории пучков.

Лемма 3. Если \mathcal{F} — дискретный $O(d)$ -bundle над связным графом и $r := \max_{\gamma_{v \rightarrow u}, \gamma'_{v \rightarrow u}} \|P_{\gamma_{v \rightarrow u}} - P_{\gamma'_{v \rightarrow u}}\|$, то $\lambda_{\mathcal{F}}^0 \leq r^2/2$.

Следствие этого результата: всегда существует нетривиальное гармоническое пространство (т.е. $\lambda_{\mathcal{F}}^0 = 0$), если транспортные отображения, порождённые ортогональным пучком, не зависят от пути (т.е. $r = 0$).

Обучение пучков из данных

Одним из ключевых практических вкладов статьи является метод обучения пучков из данных. Вместо использования заранее заданных пучков, авторы предлагают обучать restriction maps с помощью параметрической функции Φ :

$$\mathcal{F}_{v \trianglelefteq e} = \Phi(x_v, x_u)$$

где x_v и x_u — признаки вершин v и u , инцидентных ребру e .

Предложение 18. Пусть $G = (V, E)$ — конечный граф с признаками X . Тогда, если $(x_v, x_u) \neq (x_w, x_z)$ для любых $(v, u) \neq (w, z) \in E$ и Φ — MLP с достаточной ёмкостью, то Φ может обучиться любому пучку $(F; G)$.

Этот результат формально мотивирует обучение пучка на каждом слое, поскольку модель может научиться различать больше вершин после каждого шага агрегации.

Neural Sheaf Diffusion Model

Архитектура модели

Авторы предлагают модель диффузионного типа:

$$\dot{X}(t) = -\sigma(\Delta_{\mathcal{F}(t)}(I_n \otimes W_1)X(t)W_2)$$

где $\Delta_{\mathcal{F}(t)}$ — sheaf Laplacian пучка $(G, \mathcal{F}(t))$, который эволюционирует во времени. Эволюция структуры пучка описывается обучаемой функцией данных:

$$(G, \mathcal{F}(t)) = g(G, X(t); \theta)$$

Это позволяет модели использовать последние доступные признаки для манипулирования подлежащей геометрией графа и, неявно, поведением процесса диффузии.

В экспериментах авторы фокусируются на временно дискретизированной версии этой модели:

$$X_{t+1} = X_t - \sigma(\Delta_{\mathcal{F}(t)}(I \otimes W_1^t)X_t W_2^t)$$

Эта модель отличается от классической SCN модели двумя основными способами:

1. Используется обучение пучка, что делает модель применимой к любым реальным графовым датасетам
2. Используется остаточная параметризация дискретизированного процесса диффузии, что эмпирически улучшает производительность

Типы SheafLearner'ов

Для обучения restriction maps используются различные типы функций Φ :

- **LocalConcatSheafLearner:** Конкатенирует признаки соседних вершин и пропускает через линейный слой с активацией
- **AttentionSheafLearner:** Использует механизм внимания для определения весов отображений
- **QuadraticFormSheafLearner:** Использует квадратичную форму для обучения отображений

В зависимости от типа матриц, которые они обучают, различают:

- **Диагональные отображения:** Меньше параметров на ребро, упрощённая структура sheaf Laplacian. Недостаток: измерения стеблей взаимодействуют только через умножение слева на W_1 .
- **Ортогональные отображения:** Могут смешивать различные измерения стеблей, ортогональность предотвращает переобучение, лучше понимаемые теоретические свойства. В модели ортогональные матрицы строятся из композиции отражений Хаусхолдера.
- **Общие отображения:** Максимальная гибкость, но опасность переобучения. Sheaf Laplacian сложнее нормализовать численно, требуется использование SVD.

Вычислительная сложность

GCN из уравнения (2) имеет сложность $O(nc^2 + mc)$, где c — число каналов, а m — число рёбер. Предположим, что модель диффузии на пучках имеет размерность стебля d и f каналов таких, что $d \times f = c$ (т.е. тот же размер представления).

- При использовании диагональных отображений сложность составляет $O(nc^2 + mdc)$
- При использовании ортогональных или общих матриц сложность становится $O(n(c^2 + d^3) + m(cd^2 + d^3))$

На практике авторы используют $1 \leq d \leq 5$, что эффективно приводит к постоянному накладному расходу по сравнению с GCN.

Экспериментальные результаты

Синтетические эксперименты

Авторы рассматривают простую установку: связный двудольный граф с равными по размеру долями. Признаки выбираются из двух перекрывающихся изотропных гауссовых распределений, чтобы сделать классы линейно неразделимыми в момент инициализации.

Из Предложения 9 известно, что модели диффузии, использующие симметричные restriction maps, не могут разделить классы в пределе, в то время как процесс диффузии, использующий отрицательные транспортные отображения, может это сделать.

Результаты показывают, что:

- При времени диффузии ноль (т.е. без диффузии) линейный классификатор не может разделить классы
- На более поздних временах процесс диффузии, использующий симметричные отображения, не может идеально подогнать данные
- С более общей диффузией на пучках, по мере увеличения времени и приближения сигнала к гармоническому пространству, модель становится лучше, и признаки становятся линейно разделимыми
- В соответствии с Предложением 10, модель обучается отрицательному транспортному отображению для всех рёбер, что показывает, что модель способна избежать oversmoothing

Эксперименты на реальных данных

Авторы testируют свои модели на множестве реальных датасетов с коэффициентом гомофилии рёбер h от $h = 0.11$ (очень гетерофильный) до $h = 0.81$ (очень гомофильтый). Это даёт представление о том, как модель работает на всём этом спектре.

Результаты показывают, что модели NSD (Neural Sheaf Diffusion):

- Занимают первое место в 5 из 6 бенчмарков с высокой гетерофилией ($h < 0.3$)
- Показывают сильные результаты на гомофильтых графах, находясь в пределах примерно 1% от лучшей модели
- В целом, модели NSD входят в тройку лучших на 8 из 9 датасетов
- Модель O(d)-bundle диффузии показывает лучшие результаты в целом, подтверждая интуицию, что она может лучше избегать переобучения, одновременно трансформируя векторы достаточно сложными способами

Обоснование актуальности для задачи мультимодальных пучков

Почему пучки подходят для мультимодальных данных

Пучки на графах предоставляют естественный механизм для работы с мультимодальными данными:

- **Гибкость представления:** Каждая вершина может иметь своё векторное пространство, соответствующее типу модальности (видео, текст, заголовок, речь)
- **Структурированная передача информации:** Линейные отображения на рёбрах позволяют контролировать преобразовывать признаки между модальностями
- **Естественное моделирование взаимодействий:** Разные типы контента могут взаимодействовать через общие пучковые отображения

В контексте динамики мнений (opinion dynamics), x_v представляет «частное мнение» вершины v (например, эмбеддинг видео), а $\mathcal{F}_{v \leq e} x_v$ выражает, как это мнение проявляется публично в «дискурсивном пространстве», образованном $\mathcal{F}(e)$ (общее пространство для взаимодействия модальностей).

Как Neural Sheaf Diffusion помогает решить задачу

Механизм диффузии на пучках позволяет агрегировать информацию от соседних вершин с учётом структуры пучка. Модель может научиться оптимальным способом преобразования признаков между модальностями в процессе обучения. Подход работает на графах произвольной структуры, что важно для рекомендательных систем.

Конкретные преимущества для рекомендательных систем:

- **Сохранение семантики:** Пучковая структура сохраняет семантические отношения между элементами контента
- **Адаптивность:** Модель может адаптироваться к различным типам взаимодействий между модальностями
- **Интерпретируемость:** Линейные отображения пучка могут быть интерпретированы как преобразования между модальностями
- **Решение проблемы oversmoothing:** Пучковая структура позволяет контролировать процесс сглаживания признаков
- **Работа с гетерофилией:** Подход эффективен даже когда соседние вершины имеют разные типы контента

Связь с задачей курсовой работы

Методы из статьи Neural Sheaf Diffusion будут использованы в реализации модели для объединения мультимодальных эмбеддингов. Компоненты статьи будут адаптированы для мультимодального случая, где разные типы контента (видео, текст, заголовки, речь) взаимодействуют через общие пучковые отображения.

Ожидается, что такой подход превзойдёт простые методы агрегации за счёт:

- Сохранения структуры данных и семантических связей
- Контролируемого преобразования признаков между модальностями

- Избежания проблем oversmoothing и гетерофилии
- Возможности интерпретации преобразований между модальностями

Практические приложения и направления исследований

Применение к рекомендательным системам

В контексте рекомендательных систем VK (например, VK Видео) подход на основе пучков может быть использован для:

- Объединения эмбеддингов различных модальностей (видео, текст, заголовки, речь) в единое представление
- Улучшения качества рекомендаций за счёт учёта семантических связей между различными типами контента
- Построения более репрезентативных мультимодальных эмбеддингов для задач ранжирования

Направления дальнейших исследований

Статья открывает несколько направлений для дальнейших исследований:

- **Расширение до мультиmodalного случая:** Адаптация подхода для работы с несколькими модальностями одновременно, где разные типы контента взаимодействуют через общие пучковые отображения
- **Высшие порядки sheaf Laplacians:** Использование sheaf Laplacians высших порядков, которые работают с тензорами высших порядков и могут кодировать важную информацию о подлежащих симметриях в данных
- **Применение к другим задачам:** Расширение подхода к другим типам данных, таким как решётки и их связанные sheaf Laplacians
- **Категориальная интерпретация:** Использование категориальной интерпретации пучков для расширения моделей к более экзотическим типам данных

Заключение и значимость

Работа Боднара и соавторов представляет фундаментальный вклад в область графовых нейронных сетей, объединяя методы алгебраической топологии (конкретно, теорию пучков) с современными подходами к обучению на графах.

Ключевые достижения работы:

- Теоретическое обоснование связи между «геометрией» графа и проблемами гетерофилии и oversmoothing в GNN
- Демонстрация того, что при правильной структуре пучка можно избежать этих проблем
- Доказательство того, что дискретизированные параметрические процессы диффузии имеют больший контроль над асимптотическим поведением, чем классические GCN

- Практический метод обучения пучков из данных, делающий подход применимым на реальных датасетах
- Конкурентоспособные результаты на широком спектре датасетов, от сильно гетеро- фильных до гомофильтных графов

С практической точки зрения, подход на основе пучков особенно актуален для задач объединения мультимодальных эмбеддингов в рекомендательных системах. Гибкость представления, структурированная передача информации и возможность интерпретации делают его привлекательным решением для задач, где необходимо учитывать различные типы контента и их взаимодействие.

Статья предоставляет новые связи между GNN и алгебраической топологией и будет интересна обеим областям. Для задачи курсовой работы по мультимодальным пучкам на графах эта статья служит теоретической основой и практическим руководством для реализации подхода к объединению эмбеддингов различных модальностей.

Математические результаты и доказательства

Теоретические результаты о выразительной силе

Авторы доказывают ряд важных результатов о способности процессов диффузии на пучках к линейному разделению классов в пределе бесконечного времени.

Предложение 8. Пусть \mathcal{G} — множество связных графов $G = (V, E)$ с двумя классами $A, B \subset V$ таких, что для каждой $v \in A$ существует $u \in B$ и ребро $(v, u) \in E$. Тогда H_{sym}^1 имеет силу линейного разделения над \mathcal{G} .

Предложение 9. Пусть \mathcal{G} — множество связных двудольных графов $G = (A, B, E)$ с разбиениями A, B , образующими два класса, и $|A| = |B|$. Тогда H_{sym}^1 не может линейно разделить классы любого графа в \mathcal{G} для любых начальных условий $X(0) \in \mathbb{R}^{n \times f}$.

Предложение 10. Пусть \mathcal{G} содержит все связные графы $G = (V, E)$ с двумя классами $A, B \subseteq V$. Рассмотрим пучок $(F; G) \in H^1$ с $\mathcal{F}_{v \leq e} = -\alpha_e$ если $v \in A$ и $\mathcal{F}_{u \leq e} = \alpha_e$ если $u \in B$ с $\alpha_e > 0$ для всех $e \in E$. Тогда диффузия, индуцированная $(F; G)$, может линейно разделить классы G для почти всех начальных условий, и H^1 имеет силу линейного разделения над \mathcal{G} .

Эти результаты показывают, что при правильной структуре пучка процесс диффузии может достичь линейного разделения классов даже в сложных случаях, таких как двудольные графы, где классические GNN терпят неудачу.

Контроль над асимптотическим поведением

Теорема 15. Для $(F, G) \in H^1$, дискретизированный параметрический процесс диффузии имеет больший контроль над своим асимптотическим поведением, чем GCN. В частности, семейство SCN может, если необходимо, избежать ядра Laplacian.

Предложение 17. Для любого связного графа G и $\varepsilon > 0$ существует пучок $(G, F) \notin H_{\text{sym}}^d$ с $\|W_1\|_2 < \varepsilon$ и вектор признаков x такой, что $E_F((I \otimes W_1)x) > E_F(x)$.

Это важный результат, показывающий, что не только диффузия на пучках более выразительна, чем тепловая диффузия, но и SCN более выразительны, чем GCN, в том смысле, что они, как правило, не ограничены уменьшением энергии Дирихле при использовании весов с малой нормой.

Спектральный анализ и неравенство Cheeger

Авторы предоставляют новое неравенство типа Cheeger для спектрального зазора sheaf Laplacian:

Предложение 3. Если F — дискретный $O(d)$ -bundle над связным графом и $r := \max_{\gamma_{v \rightarrow u}, \gamma'_{v \rightarrow u}} \|P_{\gamma_{v \rightarrow u}} - P_{\gamma'_{v \rightarrow u}}\|$, то $\lambda_F^0 \leq r^2/2$.

Этот результат связывает спектральный зазор sheaf Laplacian с зависимостью транспортных отображений от пути, что является ключевым для понимания свойств гармонического пространства.

Детальное описание экспериментальных результатов

Синтетические эксперименты

В синтетических экспериментах авторы рассматривают связный двудольный граф с равными по размеру долями. Признаки выбираются из двух перекрывающихся изотропных гауссовых распределений, чтобы сделать классы линейно неразделимыми в момент инициализации.

Результаты показывают:

- При времени диффузии ноль (без диффузии) линейный классификатор не может разделить классы
- На более поздних временах процесс диффузии, использующий симметричные отображения, не может идеально подогнать данные
- С более общей диффузией на пучках, по мере увеличения времени и приближения сигнала к гармоническому пространству, модель становится лучше, и признаки становятся линейно разделимыми
- Гистограмма всех транспортных (скалярных) отображений показывает, что модель обучается отрицательному транспортному отображению для всех рёбер, в соответствии с Предложением 10

Результаты на реальных датасетах

Авторы testируют модели на 9 датасетах с коэффициентом гомофилии рёбер от $h = 0.11$ (очень гетерофильный) до $h = 0.81$ (очень гомофильтый):

- **Гетерофильные датасеты ($h < 0.3$):** Texas ($h = 0.11$), Wisconsin ($h = 0.21$), Film ($h = 0.22$), Squirrel ($h = 0.22$), Chameleon ($h = 0.22$), Cornell ($h = 0.23$)
- **Гомофильтые датасеты ($h > 0.7$):** Citeseer ($h = 0.74$), Pubmed ($h = 0.80$), Cora ($h = 0.81$)

Ключевые результаты:

- Модели NSD занимают первое место в 5 из 6 бенчмарков с высокой гетерофилией
- На оставшемся датасете (Chameleon) модели NSD занимают второе место
- На гомофильтых графах NSD показывает сильные результаты, находясь в пределах примерно 1% от лучшей модели
- В целом, модели NSD входят в тройку лучших на 8 из 9 датасетов
- Модель $O(d)$ -bundle диффузии показывает лучшие результаты в целом, подтверждая интуицию, что она может лучше избегать переобучения, одновременно трансформируя векторы достаточно сложными способами
- Модель, обучающая диагональные отображения, также показывает сильные результаты, несмотря на более простую функциональную форму Laplacian

Сравнение с baseline-методами

Модели NSD сравниваются с широким набором GNN моделей, которые можно разделить на три категории:

1. **Классические модели:** GCN, GAT, GraphSAGE
2. **Модели, специально разработанные для гетерофильных настроек:** GGCN, Geom-GCN, H2GCN, GPRGNN, FAGCN, MixHop
3. **Модели, решающие проблему oversmoothing:** GCNII, PairNorm

Результаты показывают, что подход на основе пучков превосходит или конкурирует с этими методами на широком спектре датасетов, демонстрируя универсальность подхода.

Связанные работы и обсуждение

Sheaf Neural Networks и обучение пучков

Sheaf Neural Networks с заранее заданным sheaf Laplacian были первоначально введены в игрушечной экспериментальной установке. В отличие от предыдущих работ, авторы предоставляют обширный теоретический анализ, обосновывающий использование пучков в Graph ML, и впервые изучают, как пучки могут быть обучены из данных с использованием нейронных сетей.

Последующие работы также экспериментировали с выводом connection Laplacian непосредственно из данных на этапе предобработки, комбинированием пучков с вниманием и проектированием моделей на основе волнового уравнения на пучках.

Гетерофилия и Oversmoothing

Хотя хорошие эмпирические конструкции, совместно решающие эти две проблемы, были предложены ранее, работа Yan et al. [72] — единственная другая работа, теоретически связывающая эти два явления. Однако их анализ очень отличается по методам и предположениям, и поэтому их результаты полностью ортогональны.

Авторы используют новый набор математических инструментов из теории клеточных пучков, что привносит новый язык и новые инструменты для анализа этих проблем.

Категориальная теория и GNN

С точки зрения категориальной теории, клеточные пучки являются функтором из категорий, описывающей структуру инцидентности графа, в категорию, описывающую данные, живущие на графике. Эта общность может быть использована для расширения моделей к более экзотическим типам данных, таким как решётки и их связанные sheaf Laplacians.

План реализации для курсовой работы

Этап 1: Реализация базового алгоритма для одной модальности

На первом этапе необходимо реализовать базовый алгоритм обучения пучка на графике для работы с одной модальностью. Это позволит получить практический опыт работы с концепцией пучков и проверить корректность реализации.

Основные компоненты:

1. Реализация структуры данных для представления пучка (стебли вершин и рёбер, restriction maps)
2. Реализация SheafLearner для обучения restriction maps (начнём с LocalConcatSheafLearner)
3. Реализация процесса диффузии на пучке
4. Реализация sheaf Laplacian и его нормализованной версии
5. Базовый цикл обучения с обратным распространением ошибки

Используемые инструменты:

- PyTorch для реализации нейронных сетей
- PyTorch Geometric для работы с графами
- NumPy для численных вычислений

Этап 2: Расширение до мультимодального случая

На втором этапе модель расширяется для работы с несколькими модальностями одновременно.

Ключевые модификации:

1. Определение различных типов стеблей для разных модальностей (видео, текст, заголовки, речь)
2. Реализация общих пучковых отображений для преобразования между модальностями
3. Адаптация процесса диффузии для работы с мультимодальными признаками
4. Механизм объединения мультимодальных эмбеддингов после диффузии

Этап 3: Эксперименты и оценка

На третьем этапе проводятся эксперименты на открытых рекомендательных датасетах.

Экспериментальная установка:

- Выбор датасетов с мультимодальными данными (например, MovieLens с метаданными фильмов)
- Предобработка данных: извлечение эмбеддингов для различных модальностей
- Построение графа взаимодействий (пользователь-контент, контент-контент)
- Разделение данных на обучающую, валидационную и тестовую выборки

Метрики качества:

- Recall@K — доля релевантных элементов среди первых K рекомендаций
- NDCG (Normalized Discounted Cumulative Gain) — нормализованный дисконтированный накопительный выигрыш
- MRR (Mean Reciprocal Rank) — средний обратный ранг

Baseline-методы для сравнения:

- Простые методы агрегации (конкатенация, усреднение эмбеддингов)
- Классические GNN (GCN, GAT) без учёта мультимодальности
- Другие методы объединения мультимодальных эмбеддингов (early fusion, late fusion, cross-modal attention)

Заключение и значимость

Работа Боднара и соавторов представляет фундаментальный вклад в область графовых нейронных сетей, объединяя методы алгебраической топологии с современными подходами к обучению на графах. Теоретические результаты о связи между «геометрией» графа и проблемами гетерофилии и oversmoothing, а также практический метод обучения пучков из данных делают эту работу особенно ценной.

Для задачи курсовой работы по мультимодальным пучкам на графах эта статья служит:

- **Теоретической основой** для понимания того, как пучки могут быть использованы для работы с различными типами данных
- **Практическим руководством** для реализации подхода к объединению эмбеддингов различных модальностей
- **Источником вдохновения** для адаптации методов к специфическим задачам рекомендательных систем

Ожидается, что применение подходов из статьи к задаче объединения мультимодальных эмбеддингов позволит создать более эффективные модели рекомендательных систем, способные учитывать семантические связи между различными типами контента и избегать проблем, характерных для классических GNN.