«Московский физико-технический институт
(национальный исследовательский университет)»
Физтех-школа прикладной математики и информатики
Кафедра «Интеллектуальные системы»
Бишук Антон Юрьевич

Контролируемая генерация графов

03.03.01 — Прикладные математика и физика

Выпускная квалификационная работа магистра

Научный руководитель:
к.ф.-м.н.
Зухба Анастасия Викторовна

Москва
2023 г.

Содержание

Аннотация

Данная работа описывает новый метод генерации графов, который использует разделение статистических характеристик графа на две группы. Первая группа, называемая «простыми признаками», может быть вычислена эффективными детерминированными алгоритмами со сложностью не более квадратичной от числа вершин. Вторая группа статистических характеристик генерируется в скрытом пространстве и затем используется для формирования матрицы смежности графа. Этот подход позволяет генерировать графы с точно заданными статистическими характеристиками, при этом сохраняя их разнообразие. Более того, данный метод может быть применен для генерации графов, имеющих сходную структуру с исходным графом, что особенно полезно при работе с графами, описывающими контакты между людьми, например, граф контактов сотрудников офиса.

1 Введение

Использование графов в качестве источников данных в задачах машинного обучения становится все более популярным, но недостаточно развито в силу отсутствия достаточно больших наборов данных для обучения и тестирования моделей. Для решения этой проблемы часто используются задачи, основанные на единственном графе, либо синтетические генераторы графов(!! https://arxiv.org/pdf/1903.03894.pdf). Однако эти методы могут быть не совершенными и не всегда коррелирующими с имеющимися данными.

Кроме того, часто возникает необходимость в графах, имеющих сходное распределение с имеющимся графом. Это важно, например, в случае графа контактов фиксированного сообщества, где возможно сгенерировать ряд похожих на него графов. Для этой задачи традиционно используются генеративные модели, однако существующие методы могут ограничивать возможности генерации и не учитывать интуитивно понятные характеристики, такие как число клик или число контактов.

Для такого рода задач традиционно используют генеративные модели. Для этого подходят как простые модели, такие как GVAE(!! "Variational Graph Auto-Encoders"Kipf & Welling, 2016), так и диффузионные (!! https://arxiv.org/pdf/2106.10934.pdf).

Однако существующие методы так или иначе сильно ограничивают возможности генерации. Они фокусируются на реконструкции исходного графа, а также не способны учитывать в генерации интуитивно понятные характеристики(например время общения человека в офисе ограничено и есть ограничения на время и число контактов). Мы же предлагаем использовать в качестве ограничения на генерацию не сколько реконстуркцию, сколько заранее выбранные глобальные статистики графа (например число ребер, вершин, кластерное число и так далее). Тем самым мы можем генерировать графы с заранее выбранными статистиками, но имеющими схожее распределение с исходным графом.

Кроме того, этот метод может быть использован для поиска сложных(пояснить) статистик графа, таких как центральность смежности (ссылка), которые до сих пор остаются нерешенными на достаточно высоком уровне. Подробнее об использовании подхода в других задачах мы расскажем в разделе с методом.

В практической части работы приведены эксперименты, иллюстрирующие превосходство нашего метода в генерации графов с заданными статистиками.

2 Обзор литературы

Генерация графов - это важный и активно развивающийся направление в машинном обучении, который находит применение во многих областях, включая биоинформатику, социальные науки, физику и многие другие. Генерация графов может помочь решать разнообразные задачи, такие как поиск наиболее важных узлов в графе, классификация графов, прогнозирование свойств графов и т.д. Существует множество методов генерации графов, таких как генеративно-состязательные сети, вариационные автоэнкодеры, нормализующие потоки и многие другие.

2.1 Генеративные модели

Одной из ключевых работ в области глубокого обучения и вероятностного моделирования является Статья "Auto-Encoding Variational Bayes"(D. Kingma and M. Welling, 2013), представляющей собой первое упоминание модели вариационных автокодировщиков (VAE).

В статье авторы представляют подход к генеративному моделированию данных, который позволяет моделировать сложные распределения и обеспечивает более эффективное обучение в сравнении с классическими методами. Они предлагают использовать нейронную сеть в качестве генеративной модели, которая будет преобразовывать входные данные в латентное пространство, а затем обратно декодировать из латентного пространства в исходное. Авторы представляют новый функционал ошибки для обучения модели, основанный на вариационном выводе. Этот функционал позволяет обучать модель вариационным методом и позволяет получать оценки правдоподобия для новых сгенерированных данных. Статья является ключевой в развитии вероятностного моделирования в глубоком обучении и открыла новые возможности для генеративного моделирования данных, включая генерацию графов при помощи вариационных автокодировщиков.

Наш метод основывается на использование дополнительной информации при генерации. Первое упоминание такой идеи было в статье "Learning Structured Output Representation using Deep Conditional Generative Models"от David J. Rezende, Shakir Mohamed и Daan Wierstra, где, впервые была представлена модель Conditional Variational Autoencoders (CVAE).

CVAE - это модификация VAE, которая может генерировать данные с заданными условиями. В стандартном VAE модель генерирует данные на основе скрытого

пространства, которое не зависит от каких-либо внешних переменных. В CVAE модель использует дополнительную информацию для генерации данных. Авторы в статья показывают, как CVAE может быть использован для генерации изображений с заданными свойствами. Они используют MNIST-датасет для генерации цифр с определенными свойствами, такими как цвет и положение цифры на изображении. Кроме того описывается, как CVAE может быть использован для классификации изображений. Авторы применяют CVAE к задаче классификации CIFAR-10, показывая, что CVAE может значительно улучшить точность классификации. С тех пор было предложено множество модификаций CVAE, таких как AC-GAN (Auxiliary Classifier GAN) и InfoGAN (Information Maximizing GAN), которые используют подобные идеи для генерации изображений с более сложными свойствами.

Идея нашего метода основывалась на выводах, что подмешенная дополнительная информация помогает при генерации данных. Однако графы это специфические данные, которые обладают своими свойствами, которые можно использовать. Нам нет нужды добывать дополнительную информацию извне, поскольку мы можем выделить ее из наших же данных. С этой точки зрения наш метод можно назвать Self-Conditional VAE.

!! НУжны картинки с VAE и CVAE (Нарисовать самому)

2.2 Методы обработки графов

!! СЮДА НУЖНА КАРТИНКА (https://tkipf.github.io/graph-convolutional-networks/)

Самым популярным методом обработки информации в графах является Graph Convolutional Networks (GCNs) - это класс нейросетевых архитектур, которые применяются для анализа данных на графах. Они представляют собой расширение сверточных нейронных сетей (CNN) для графовой структуры данных.

Метод был представлен в Semi-Supervised Classification with Graph Convolutional Networks. Авторы отмечают, что прежде чем GCN были предложены, для анализа графов применялись методы, такие как Graph Laplacian и DeepWalk. Однако, эти методы имеют недостатки, связанные с ограничениями по сложности моделей и слишком большим размером представлений графов соответственно.

GCN объединяет свойства CNN и графовых моделей, которые позволяют применять сверточные операции непосредственно на графовых структурах. В частности,

они вводят операцию Graph Convolution, которая агрегирует информацию соседних узлов графа и создает новое представление для текущего узла. GCN также предоставляет подробное описание архитектуры, включая алгоритм обучения и вычислительные сложности. Авторы привели примеры использования GCN для классификации вершин в графах и для задачи сегментации изображений, а также обсудили проблемы переобучение и предложили методы регуляризации модели.

В целом, GCN является мощным инструментом для анализа графов и может быть применен в различных областях, включая социальные сети, биоинформатику, рекомендательные системы и многие другие.

2.3 Генерация графов при помощи VAE

Генерация графов - это задача, которая привлекает внимание многих исследователей в области компьютерных наук и машинного обучения. В последние годы было предложено множество методов для генерации графов, используя различные подходы.

Variational autoencoders (VAE) - это класс моделей глубокого обучения, которые могут быть использованы для генерации графов. Основная идея заключается в том, чтобы скрыть некоторую структуру, называемую латентным пространством, в котором графы могут быть сгенерированы. В этом обзоре мы рассмотрим несколько статей, в которых были использованы VAE для генерации графов в хронологическом порядке.

Одной из первых статей, посвященных генерации графов при помощи VAE является "Variational Graph Auto-Encoders"(Kipf & Welling, 2016). Авторы предложили использовать VAE для изучения скрытого распределения графов и генерации новых графов путем сэмплирования из этого распределения. В работе был предложен графический энкодер, который принимает матрицу смежности графа и преобразует ее в скрытый вектор. В декодере происходит обратное преобразование скрытого вектора в матрицу смежности.

Развитие этой идеи произошло в статье "GraphVAE: Towards Generation of Small Graphs Using Variational Autoencoders"(Simonovsky Komodakis, 2018), где VAE использовалась для генерации маленьких графов. Авторы предложили модифицированный VAE, который использует два различных энкодера, один для работы с вершинами графа, а другой - для работы со связями между вершинами. Также в статье

была предложена функция потерь, которая позволяет измерять качество генерации графов по нескольким метрикам.

Далее графовые VAE модели начали использовать во многих профильных областях. Так в статье "Junction Tree Variational Autoencoder for Molecular Graph Generation"(Jin et al., 2019) использовалась VAE для генерации молекулярных графов. В статье был предложен способ использования деревьев соединений для представления молекул, и была разработана новая функция потерь, которая позволяет управлять генерацией молекулярных графов.

В целом, использование VAE для генерации графов является перспективной идеей, и в последние годы было предложено множество модификаций VAE для решения этой задачи. Также были предложены различные функции потерь и метрики для оценки качества генерации графов.

3 Основные понятия

В этом разделе мы введем понятия, которые далее будем использовать.

3.1 Используемые понятия

Определение 3.1. *Простые признаки (или же простые статистики) в нашем методе - это признаки графа, которые могут быть выражены через простые статистики, такие как степень вершин, число треугольников в графе, средняя длина пути, диаметр графа и т.д. Они являются более простыми и прямолинейными характеристиками графа, которые могут быть легко вычислены и использованы для генерации новых графов.*

Определение 3.2. *Сложными признаками (или статистиками) мы называем все, что не является простыми признаками, а также от простых признаков не зависит.*

Определение 3.3. *Смешанными признаками назовем какую-либо линейную комбинацию сложных и простых признаков. К случае обычной генерации графов в латентном пространстве будут находиться как раз смешанных статистики.*

Определение 3.4. *kl дивергения ??????????????????????????*

3.2 Статистики графов

В качестве простых признаков графа нами были выбраны следующие характеристики:

- Размерные показатели:
  - Число ребер
  - Число вершин
- Вершины специального вида:
  - Изолированные вершины – вершины без единого ребра
  - Висячие вершины – вершины с одним ребром
  - Промежуточные вершины – вершины с двумя ребрами
  - Вершины, связанные с каждой вершиной графа
- Статистики на степенях вершин:
  - Максимальная степень вершины
  - Средняя степень вершины
  - Медианная степень вершины
  - Модальная степень вершины
  - Стандартное отклонение степеней вершин в графе
- Гистограмма степеней вершин графа (здесь μ – средняя степень вершин в графе, σ – среднеквадратичное отклонение степеней вершин в графе):
  - Доля вершин со степенью на интервале $(\mu - \sigma, \mu)$.
  - Доля вершин со степенью на интервале $(\mu, \mu + \sigma)$.
  - Доля вершин со степенью на интервале $(\mu - 2\sigma, \mu - \sigma)$.
  - Доля вершин со степенью на интервале $(\mu + \sigma, \mu + 2\sigma)$.
  - Доля вершин со степенью на интервале $(\mu - 3\sigma, \mu - 2\sigma)$.
  - Доля вершин со степенью на интервале $(\mu + 2\sigma, \mu + 3\sigma)$.
- Размер наибольшей клики в графе

- Коэффициент кластеризации
- Статистики связанные с циклами в графе:
  - Наличие цикла.
  - Размер базиса циклов.
  - Максимальный элемент в цикле базисов.
  - Минимальный элемент в цикле базисов.

4 Предлагаемый метод

4.1 Описание метода

Новый метод генерации графов, основанный на модели вариационного автоэнкодера (VAE), является улучшенной версией обычного VAE. Основное преимущество нового метода заключается в том, что он позволяет явно контролировать генерацию графов с заданными свойствами, что делает его более удобным и гибким по сравнению с обычным VAE.

В нашем методе мы разделяем матрицу признаков графа на «простые статистики» и «сложные статистики». Простые статистики представляют собой свойства графа, которые можно вычислить эффективными детерминированными алгоритмами с асимптотикой не выше квадратичной от числа вершин, например, степени вершин или количество ребер. Сложные статистики, напротив, представляют собой более сложные свойства графа, которые не могут быть вычислены с помощью простых алгоритмов, например, максимальная центральность смежности в графе или длина максимального цикла в графе.

Мы избавляемся от простых признаков в матрице смешанных признаков и генерируем математическое ожидание и дисперсию для сложных признаков. Затем мы генерируем скрытую матрицу сложных статистик и добавляем к ней простые статистики, чтобы получить итоговую матрицу признаков графа. Таким образом, мы можем явно указать, какими свойствами должен обладать сгенерированный граф, что делает наш метод более удобным и гибким по сравнению с обычным VAE.

Кроме того, наш метод позволяет генерировать графы с более точными фиксированными статистиками при большом разнообразии сгенерированных примеров. Это

достигается за счет контроля за процессом генерации, а также за счет улучшенного скрытого пространства. Таким образом, наш метод имеет больший потенциал для использования в различных областях, где требуется генерация графов с заданными свойствами.

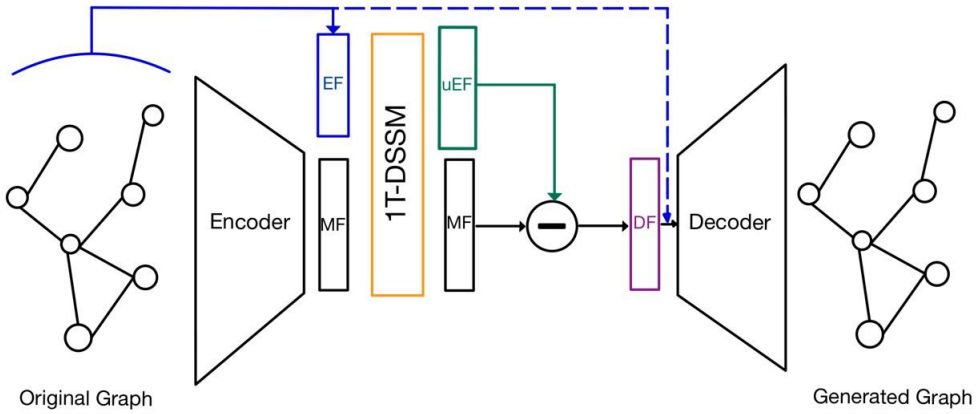

Рис. 1: Схема предложенного метода. !!Перерисовать

4.2 Теоремы

Стандартный метод формирования скрытого представления в VAE это трюк репараметризации. Благодаря ему, мы гарантируем, что скрытое представление принадлежать заданному семейству распределений (чаще всего нормальному). Таким образов мы гарантируем, что смешанные статистики графа будут иметь нормальное распределение.

Введем понятие пространство элементарных статистик графа. Если мы рассматриваем графы с фиксированным числом вершин n, то можем говорить, что один из таких графов представим в виде радиус-вектора в пространстве элементарных статистик. Тогда в такой постановке простыми статистиками будем называть такие вектора, которые ортогональны между друг другом и кроме того находятся в маломерном подпространстве. Смешанными же признаками будут являться любые вектора в таком пространстве.

Простыми статистиками графа назовем посчитанные детерменированными алгоритмами численные характеристики графа. Будем считать такие вектора также нормальными.

Для теоретического обоснования предложенного метода сформулируем и докажем ряд теорем.

Теорема 4.1 (Бишук 2023). *Пусть задано вероятностное пространство, состоящее из характеристик графа. Рассмотрим нормальный вектор, принадлежащий подпространству данного вероятностного пространства - будем называть его вектором смешанных статистик. Если заданы простые статистики графа, которые являются компонентами вектора смешанных статистик, то возможно этого вектор смешанных статистик разложить в сумму вектора сложных признаков и вектора простых признаков.*

Доказательство.

В качестве доказательства приведем алгоритм построения вектора сложных признаков:

Шаг 0: Начнем рассуждения для случая фиксированных графа и простых статистик. Обозначим G как граф, m – размерность вектора смешанных статистик графа G, s – размерность вектора простых статистик графа G.

Шаг 1: Составим систему из s уравнений с m неизвестными и решив ее построим линейную оболочку решений $L_s = \langle l_1, l_2, \dots l_s \rangle$.

Шаг 2: Поскольку по определению сложные статистики это все, которые не являются простыми, построим ортогональное дополнение путем составления системы уравнений, полученных из уравнения:

$$\langle L_s, x \rangle = 0.$$

Таким образом мы получим линейную оболочку сложных статистик.

Шаг 3: Теперь вспомним, что мы находимся в векторном пространстве, а вектор смешанных статистик и простых статистик – это нормальные вектора. Ортогональность в линейном пространстве говорит о независимости векторов – в векторном пространстве о независимости нормальных векторов говорит нулевая ковариация. То есть на **Шаге 2**, мы ищем подпространство сложных признаков при помощи системы уравнений:

$$\mathrm{COV}(L_s, x) = 0.$$

Таким образом мы получим вектор, который будет независимым относительно простых статистик, при этом полученных линейной комбинацией из смешанных статистик. А это по определению вектор сложных статистик.

■

Теорема 4.2 (Бишук 2023). *Для нахождения сложных статистик в смешанных признаках достаточно найти такой вектор, который получается из смешанных линейным преобразованием и не является скоррелированым с простыми статистиками.*

Доказательство.

Пусть у нас есть нормальный вектор смешанных статистик $\vec{m}$, а также нормальный вектор простых статистик $\vec{s}$ и некоторое линейное преобразование NN (в случае если размерность $\vec{m}$ и $\vec{s}$ не совпадают, то это преобразование переводит $\vec{m}$ в размерность $\vec{s}$).

Применим линейное преобразование к вектору смешанных векторов $\vec{d} = NN(\vec{m})$. Линейное преобразование нормального вектора – есть нормальный вектор, потому $\vec{d}$ – нормальный вектор.

Теперь воспользуемся теоремой, что независимость двух нормальных векторов эквивалентна их нескоррелированности. Тем самым, если линейное преобразование NN перевело вектор $\vec{m}$ в вектор, который нескоррелирован с вектором $\vec{s}$, то получившийся вектор $\vec{d}$ будет вектором сложных статистик

■

4.3 Использование в смежных областях

Разделение графов на «простые» и «сложные» признаки может быть полезным в различных областях, например:

Медицина: в медицинской диагностике графы могут представлять собой молекулы лекарственных препаратов, где простые признаки могут быть связаны с физико-химическими свойствами атомов, а сложные признаки - с их биологическими свойствами, такими как взаимодействие с рецепторами.

Социальные сети: графы социальных сетей могут быть использованы для анализа социальных взаимодействий и связей между людьми. Простые признаки могут быть связаны с такими характеристиками, как возраст, пол, местоположение, а сложные признаки - с социальным статусом, интересами и т.д.

Финансы: графы могут представлять собой финансовые потоки между компаниями. Простые признаки могут быть связаны с финансовыми показателями компаний, такими как доход, расходы, прибыль, а сложные признаки - с отношениями между компаниями, такими как поставщик-потребитель, конкуренты и т.д.

В этих областях мы можем использовать разделение графов на простые и сложные признаки, чтобы более точно моделировать свойства графов. Например, мы можем использовать простые признаки для генерации графов с определенными статистическими характеристиками, такими как размер, плотность, средний путь и т.д. А сложные признаки мы можем использовать для генерации графов, которые имеют определенные свойства, связанные с конкретной областью применения, например, для медицинских молекул - с определенными фармакологическими свойствами.

5 Вычислительный эксперимент

В этом разделе мы производим проверку предложенного метода и сравниваем его с существующими методами на реальных данных.

5.1 Датасеты

Эксперимент проводился на наборе данных Cora и Citeseer.

Датасет "Cora"("The Graph Neural Network Model") (https://arxiv.org/abs/1812.08434) - это один из наиболее часто используемых датасетов в задачах классификации и кластеризации научных статей. Он состоит из 2708 статей, разбитых на 7 категорий (биология, информатика, право, математика, медицина, физика и социология). Каждая статья представлена в виде узла графа, а связи между статьями - это ссылки между ними.

Каждая статья представлена в виде мешка слов (bag-of-words) из 1433 уникальных терминов, которые были извлечены из полного текста статей. Каждый узел графа также содержит список ссылок на другие статьи в датасете.

Данные для датасета были собраны в 2002 году из базы данных arXiv .

Citeseer (J. Srivastava, R. Cooley, M. Deshpande, and P. Tan, "Web usage mining: Discovery and applications of usage patterns from web data,"SIGKDD Explorations, vol. 1, no. 2, pp. 12–23, 2000.) - это датасет для задачи классификации статей по научным темам. Датасет содержит более 3300 статей из компьютерных наук и связанных с

ними областей, таких как базы данных, информационный поиск и машинное обучение. Каждая статья представляет собой узел в графе, а ссылки на другие статьи формируют ребра. В датасете всего 6 классов научных тем: базы данных, интеллектуальная обработка информации, машинное обучение, информационный поиск, распределенные системы и робототехника.

Citeseer был создан в 1999 году и был одним из первых датасетов, который использовался для классификации научных статей. Он предоставляет возможность исследования различных методов классификации, таких как методы на основе графов, методы на основе контента и гибридные методы.

Датасет Citeseer содержит 3 327 статьи и 4 732 ссылки между ними. Каждая статья представлена в виде метаданных, включающих название, список авторов, перечень ссылок на другие статьи и аннотацию. В датасете присутствуют также информация о том, какие статьи были процитированы другими, что позволяет использовать его для задач анализа цитирования.

Данные датасета были собраны из различных источников, включая базы данных ACM, DBLP и PubMed. В оригинальной статье, описывающей датасет, авторы провели анализ структуры сети цитирования, выделили основные сообщества статей и оценили качество работы алгоритмов кластеризации и классификации на этих данных.

Для каждой статьи в датасете имеются метки классов, которые были присвоены вручную на основе тезисов и ключевых слов статей. Это делает Citeseer ценным ресурсом для задачи классификации и оценки качества методов классификации на научных данных.

5.2 Постановка задачи

Традиционно, для того, чтобы научиться генерировать новый элемент данных, необходимо научиться реконструировать объект из скрытого пространства. Это так называемое end-to-end обучение. В нашем случае мы будет предсказывать наличие и отсутсвие ребра в графе.

Формально постановка это задачи с маркированными ребрами в матрице смежности графа, которая может быть описана следующим образом.

Дано:

Граф G с n вершинами и матрицей смежности A размера (n, n), где $A_{ij} = 1$,

если ребро (i, j) существует в графе, и 0 в противном случае. Набор маркированных ребер E = (i, j) существующих в графе.

Задача:

Предсказать вероятность существования ребер между всеми парами вершин в графе, включая немаркированные ребра. Особый интерес представляет предсказание наличия маркированных ребер. Формально, задача может быть сформулирована как задача бинарной классификации для каждой пары вершин. Для каждой пары вершин i и j нужно предсказать вероятность того, что ребро (i, j) существует в графе, то есть принимает значение 1 в матрице смежности.

Модель машинного обучения обучается на обучающей выборке, которая состоит из пар вершин с маркированными и немаркированными ребрами. Она должна определить, какие признаки графа могут помочь в предсказании наличия ребер, и на основе этих признаков построить модель, которая может классифицировать каждую пару вершин в графе.

Результатом работы модели является матрица предсказанных вероятностей существования ребер между всеми парами вершин в графе, включая немаркированные ребра, а также предсказание наличия маркированных ребер.

5.3 Процесс обучения и тестирования

Обучение происходило на популярных датасетах Cite, Citeseer.

Матрица смежности предварительно нормируется по следующему алгоритму:

$$\hat{A} = D^{-1/2} A D^{-1/2}, \text{ где}$$

- A – матрица смежности графа,
- D – диагональная матрица степеней вершин.

Это необходимо чтобы не возникало взрыва и затухания градиента, а также Kipf and Welling (!! https://arxiv.org/pdf/1609.02907.pdf) показал, что это приводит к лучшей сходимости.

Наша модель получала на вход матрицу смежности, признаки вершин, а также вектор простых статистик. В энкодере, после прохода блоком GCN из получившихся признаков вершин вычитался преобразованных вектор простых статистик. Среднее MSE между признаками вершины и преобразованным вектором простых статистик

запоминался. После генерации скрытого вектора (в котором теперь должны находиться только сложные признаки), мы прибавляем преобразованных простые статистики и создаем матрицу смежности, где в ячейке матрицы находится вероятность того, что на этом месте должно быть ребро.

На каждой эпохе часть ребер случайным образом выбиралось и маркировалось, также выбиралось такое же количество случаев, когда между вершинами ребра не существовало. После этого эти ребра разделяются на ребра для обучения и для валидации.

После прохода нейронной сети и получения матрицы смежности, рассчитывался BCELoss на выбранных ребрах, кроме того прибавлялась KL-дивергенция между нашим скрытым вектором и стандартным нормальных распределением и MSELoss с этапа вычитания простых признаков.

5.4 Результаты

Поскольку результатом работы алгоритма является матрица с вероятностями, традиционно для классификации смотрят ROC-AUC и Average Precision. Для того, чтобы сравнивать насколько глобальные статистики сгенерированного графа отличаются от изначального, мы проходим со всеми возможными порогами для вероятности того, что ребро существуем и находим минимум MAE. Этот минимум и записан в таблице результатов ниже.

	Dataset	ROC-AUC	AP	MAE (global statistics)
VAE	Cite	75.18 %	**75.81%**	0.066
Our method		**76.68 %**	75.18 %	**0.046 (-30%)**
VAE	Citeseer	**82.09 %**	**79.94 %**	0.072
Our method		76.28 %	76.44 %	**0.060 (-17%)**

Таблица 1: Результаты вычислительного эксперимента по классификации наличия ребер.

Как видим, небольшое уменьшение качества классификации ребер компенсируется серьезным уменьшением ошибки в простых статистиках. Кроме того, скорее всего уменьшения качества классификации можно избежать, если использовать более сложных метод вычитания и подмешивания простых статистик.

6 Заключение

В ходе данной работы и исходя из данных результатов, можно сделать вывод, что предложенный метод имеет потенциал в улучшении генерации графов. Небольшое снижение качества реконструкции может быть компенсировано значительным улучшением в точности генерации графов по простым статистикам. Это означает, что наш метод способен генерировать графы с заданными свойствами, что может быть полезным во многих приложениях, таких как генерация молекулярных структур или сетей связей между людьми. Однако, необходимы дальнейшие исследования для того, чтобы определить, какие свойства графа могут быть улучшены с помощью данного метода и как он может быть применен в более широком контексте.

Кроме того, в будущих работах мы планируем расширить предложенный метод, чтобы использовать и другие статистики графов для более точной генерации. Мы также будем исследовать различные методы агрегации графов в векторе одной вершины, что может помочь с более точной генерацией и классификацией графов.