

Московский физико-технический институт
Факультет общей и прикладной физики

Дипломная работа

Теория случайных матриц и графы ближайших соседей

О структуре графов ближайших и следующих за ближайшим соседом

Работу выполнил:

Кристалль Александр Сергеевич, Б02-003 группа

г. Москва
2024 год

Содержание

1	Аннотация	1
2	Введение	2
2.1	Актуальность исследуемой темы	2
2.2	Теория случайных матриц	2
2.3	Матрица расстояний к-ближайших соседей	4
3	Постановка задачи	5
4	Граф ближайших соседей	6
5	Граф вторых соседей	10
6	Построение случайной матрицы и Гипотезы.	13
6.1	Построение случайной матрицы.	13
6.2	Эксперимент.	14
6.3	Гипотезы.	15
7	Заключение	17
8	Литература	18

1 Аннотация

Одной из наиболее значимых проблем исследования графовых структур является изучение структуры графов ближайших соседей, используемой в компьютерном моделировании, моделях машинного обучения и анализе социальных связей. Примером могут служить модели kNN и математические алгоритмы поиска. Метод k ближайших соседей представляется из себя один из наиболее распространенных методов кластеризации данных в метрическом пространстве. Кластеризация – это эффективный способ сокращения описания, когда вместо большого количества исходных объектов оказывается возможным рассмотреть относительно небольшое число классов и сформировать типичного представителя класса. Мотивация предпринятого исследования обусловлена трудностями, возникающими при попытке систематизации графов ближайших соседей в контексте статистического анализа свойств исходного набора объектов.

Дипломная работы посвящена исследованию структуры графов ближайших и следующих за ближайшим соседей, а также статистических закономерностей между числом компонент связности графа и числом вершин в компонентах связности. В работе представлена численная оценка приближенности полученных статистических моделей к стандартным геометрическим моделям.

2 Введение

В данной работе исследуется структура графов ближайших и следующего за ближайшим соседей, для этого следует рассмотреть несколько базовых определений и теорем, которые нам потребуются. Также будут рассмотрены различные конструкции для построения случайных матриц.

2.1 Актуальность исследуемой темы

Актуальность научных исследований по изучению структуры графов ближайших соседей (*k*-nearest neighbor graphs) проявляется в нескольких ключевых областях:

Машинное обучение и обработка данных: Графы ближайших соседей широко используются в методах машинного обучения для классификации, кластеризации и визуализации данных. Исследования в этой области могут привести к разработке более эффективных алгоритмов обработки данных и улучшению производительности моделей машинного обучения.

Социальные сети и анализ сетей: Изучение структуры графов ближайших соседей имеет применение в анализе социальных сетей, сетей связей в биологии, транспортных сетей и других областях. Это может привести к новым методам анализа и представления сложных сетевых взаимосвязей.

Компьютерное зрение и обработка изображений: В контексте компьютерного зрения, графы ближайших соседей используются для анализа и классификации изображений. Исследования в этой области могут способствовать улучшению методов распознавания образов и обработки изображений.

Информационный поиск и рекомендательные системы: Графы ближайших соседей могут быть применены в информационном поиске и создании рекомендаций, например, для поиска похожих элементов или рекомендации контента пользователям. Исследования в этой области могут способствовать улучшению алгоритмов ранжирования и персонализированных рекомендаций.

Биоинформатика и медицинская диагностика: В биоинформатике графы ближайших соседей используются для анализа геномных данных, а также в медицинской диагностике для классификации и анализа медицинских данных. Исследования в этой области могут способствовать развитию более точных и надежных методов анализа биологических данных.

Таким образом, исследования по исследованию структуры графов ближайших соседей имеют широкий спектр приложений и могут привести к разработке более эффективных методов обработки данных, классификации и анализа сложных систем.

2.2 Теория случайных матриц

Конус матриц расстояний и его геометрия. Случайные матрицы расстояний. Для определения графа ближайших соседей требуется определить матрицу инцидентности этого графа, на основе которой уже

будет построена матрица смежности. Существуют два варианта определения этой матрицы. Первым является метод выкидывания в N мерный куб и подсчёт расстояний между ними, вторым и более затратным с точки зрения сложности создания алгоритма является построение графа на основе случайной матрицы расстояний.

Рассмотрим все бесконечные матрицы вида:

$$R = \{\{r_{i,j}^s\} : r_{i,i} = 0, r_{i,j} \geq 0, r_{i,k} + r_{k,j} \geq r_{i,j}\} \quad (1)$$

матрицы, удовлетворяющие (1) будем называть матрицами расстояний.

Множество всех матриц расстояний есть выпуклый конус в векторном пространстве бесконечных вещественных матриц. Мы будем называть его конусом R матриц расстояний. Определим польское пространство как полное метрическое сепарабельное пространство. Помимо бесконечного конуса матриц расстояний, каждая из которых определяет метрику на N , можно рассмотреть конечный конус матриц расстояний R_n . Определим пространство симметричных матриц размерности N с нулями на диагонали за M_n^s и заметим, что $R_n \in M_n^s$.

Далее рассмотрим допустимые векторы и структуру конуса. Пусть r - матрица расстояний порядка n . Будем называть вектор a с размерности n допустимым вектором для фиксированной матрицы расстояний r , если дописывание его справа от матрицы как столбец и снизу как строку делает получившуюся матрицу r^a матрицей расстояний порядка $n+1$. Для допустимых векторов матрицы расстояний существует несколько полезных теорем, которые стоит привести ниже:

Лемма 1. Вектор a является допустимым для матрицы r , если выполнено следующее условие на компоненты вектора и матрицы:

$$|a_i - a_j| \leq r_{ij} \leq a_i + a_j. \quad (2)$$

За $A(r)$ примем множество всех допустимых векторов для определённой случайной матрицы r . За S_n будем обозначать симметрическую группу.

Лемма 2. Для всякой матрицы $r \in R_n$ следующие множества совпадают:

$$A(grg^{-1}) = g(A(r)), \text{ where } g \in S_n. \quad (3)$$

Способы построения случайной матрицы. Так как у матрицы есть ограничения на новые элементы (допустимые вектора) при выкидывании случайных чисел в куб (это описано в [1]), то при построении такой матрицы на основе добавления компонент к уже существующей (описано в [1] и [3]), возникает проблема сужения диапазона допустимых значений нового значения расстояния. Так как в статье 3 было показано, что такой диапазон никогда не пуст, то вариант построения такой матрицы возможен. Этот вариант построения описывается в статье 3. В работе рассмотрим более прямой и ресурсозатратный вариант построения матрицы - построение по определению, когда неравенство треугольника выполняется из метода

построения матрицы. Для этого рассмотрим N точек в d -мерном пространстве с координатами в $[0,1]$. Тогда получим случайную матрицу расстояний, элементы которой будут записаны в виде:

$$\{r_{i,j} = \sqrt{\sum r^2}\} \quad (4)$$

Таким образом мы получаем матрицу размером $N \times N$, где N -число рассматриваемых точек.

Определение графа ближайших соседей. После построения случайной матрицы одним из двух представленных способов мы получили полный граф, для этого графа рассматриваются его ближайшие соседи. В каждом столбце графа находится наименьший элемент отличный от нуля и заменяется на 1. Таким образом получается граф ближайших соседей, обладающий следующими свойствами:

- 1 с каждой вершиной связано ровно одно выходящее ребро,
- 2 нет изолированных точек,
- 3 сумма степеней вершин равна числу вершин,
- 4 каждая компонента связности графа содержит ровно один цикл длины 2.

Рассматривая данные свойства графа ближайших соседей стоит отметить, что все ГБС обладают данным свойством. Поэтому стоит задаться вопросом: для каждого ли графа обладающего свойствами ГБС существует польское пространство, в котором граф с данной структурой, описываемой свойствами 1-4 будет являться графом ближайших соседей.

2.3 Матрица расстояний к-ближайших соседей

Рассматриваем граф к-ближайших соседей. Построение этого графа аналогично построению ГБС, только при рассмотрении вершины, с которой будет связано данная вершина A , берётся не ближайшая к ней вершина B , а вершина K , которая находится на k -м месте в списке вершин, удалённых от вершины A . Также стоит отметить, что все расстояния принимаются различными, так как вероятностная мера одинаковых расстояний равна нулю, а случаи, когда сгенерированные на метрическом пространстве точки находятся на одинаковом расстоянии от какой-то конкретной составляют пренебрежимо малую величину равную погрешности вычислений координат точек.

3 Постановка задачи

Цель работы. Исследовать структуру графов ближайших и следующих за ближайшим соседней. Получить статистически закономерности и связи между количеством компонент связности в графе и числом вершин компонент связности. В рамках работы были поставлены и решены следующие задачи:

- 1 Существование метрического пространства для произвольного графа, обладающего свойствами графа ближайших соседей. Равномерная оценка размерности этого пространства.
- 2 Наличие цикла длины $N-1$ для графа с числом вершин N .
- 3 Статистическое приближение получившихся моделей к стандартным геометрическим моделям в теории случайных матриц.

Используемое оборудование. Для исследования статистических закономерностей был использован компьютер и программа на языке python 3.8. Программа доступна по ссылке [9] Для генерации случайных чисел был использован встроенный генератор случайных чисел, который использует формулу генерации случайного числа, не влияющего на итоговое распределение.

Задача доказательства:

- 1 Доказать существование цикла длины 2 и структуры графа.
- 2 Доказать теорему про МП.
- 3 Доказать что в графе вторых соседей циклы любой длины.
- 4 Обобщить что для произвольного тоже такое есть.
- 5 Рассмотреть 2 и 3 для ближайших соседей.
- 6 Рассмотреть существование пространства достаточно большой размерности в котором можно реализовать граф $2x$ ближайших соседей.

Задача экспериментов:

- 1 Описывать методику проведения эксперимента.
- 2 Классифицировать варианты построения случайной матрицы и графа ближайших соседей по типам испытаний.

4 Граф ближайших соседей

Цикл для компоненты связности. Далее при рассмотрении графа как на "картинке" (пример графа для 18 ближайших соседей). Любая компонента связности имеет ровно 1 цикл.

Теорема 1. Любая компонента связности графа ближайших соседей имеет ровно один цикл и этот цикл длины 1.

1 пусть существует связанный граф ближайших соседей с двумя циклами в одной из компонент связности этого графа. Рассмотрим цепочку вершин, которые связывают два цикла, тогда в этой цепочке найдётся вершина, из которой выходит две стрелки (в силу чётности), тогда получаем противоречие с заданием ГБС (рассматриваем вершину которая заходит в цикл, тогда в эту вершину тоже приходит направленное ребро, чтобы соединить и.т.д., но так как для цикла нужна выходящая то получается выходящая и входящая одновременно). Любые две последовательные сонаправленные направленные рёбра меняем на одну, если связано и только 1 то можно заменить всю связь, тогда должно быть 2 стрелки.

2 Рассмотрим неравенство треугольника, которое применимо для расстояний между вершинами графа. Предположим, что найден цикл, длины более 2, тогда возникает противоречие с неравенством треугольника.

Граф ближайших соседей планарен. Теорема Понтрягина — Куратовского утверждает что графы K_5 и $K_{3,3}$ (полный двудольный граф имеющий по 3 вершины в каждой доле) являются единственными минимальными непланарными графами. Эту теорему можно переписать следующим образом: граф планарен тогда и только тогда, когда он не содержит подграфа, который является подразделением K_5 или $K_{3,3}$. Рассмотрим наш граф: граф разбивается на компоненты связности. Заметим что граф планарен, если каждая из его компонент связности планарна, тогда доказательство планарности графа сводится к доказательству планарности его компонент связности. О КС мы знаем: это граф ближайших соседей (в матрице смежности такого графа мы можем переставить столбцы так, чтобы матрица стала диагональной, на диагонали которой находятся матрицы соответствующих компонент связности - отсюда мы можем сказать, что можно без потери общности рассматривать перестановку матрицы, отвечающую за преобразование матрицы в диагональную, т.е. для каждой КС рассматривать собственную матрицу смежности). Так как подразделение K_5 или $K_{3,3}$ должно содержать более 1 цикла длины 2 и более, а по доказанному КС имеет ровно один цикл, то любая КС - планарный граф, тогда весь граф ближайших соседей планарен и его можно изобразить на плоскости.

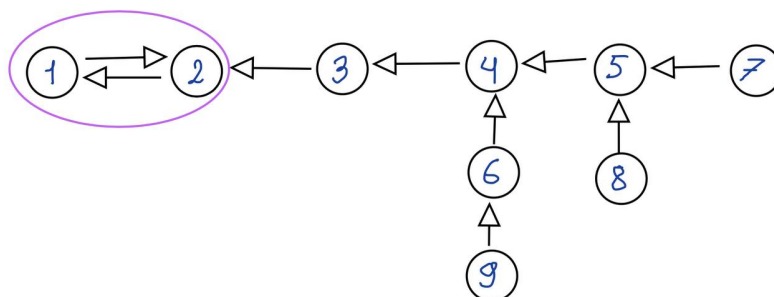


Рис. 1: Пример КС

Теорема 2: Существования польского пространства для ГБС.

Для произвольного графа, обладающего свойствами 1-4 графа ближайших соседей, существует полное метрическое сепарабельное пространство, в котором существует граф, для которого рассматриваемый является графом ближайших соседей.

Докажем существование такого пространства по индукции, перед этим зададим процесс рисования КС, основанный на ветвлении. Этот процесс нужен для поиска шага индукции. Для определённости будет считать граф связанным (обычным без направлений) и вес каждой вершины будем считать равным 1. Для начала берём вершину единственного цикла - 1 и ищем ближайшую к ней, это 2. Далее нумеруем вершины числами 3 и 4 - это очевидно. Дальше возникают две вершины, расстояние до которых от вершины 1 одинаково, таким образом получаем две вершины для которых случайным образом определяем последовательность нумерования, которую мы запоминаем и также для каждой из вершин записываем (см рис.2). Таким образом у нас получаем занумеровать вершины любой компоненты связности.

- 1 Рассмотрим цикл длины 2. Для него выполняется условие существования МП - просто линия, т.е. R^1 : помещаем одну из вершин в 0 и откладываем расстояние $(1+i/1000)$ в положительную полуось (где $i=1$), после чего ставим там вершину и получаем МП с заданными характеристиками. Метрику мы рассматриваем евклидову.
- 2 Предположим, что для КС на i -м шаге существует некоторое R^j , где j строго меньше i . Рассмотрим $k+1$ -й шаг добавления вершины.
- 3 Пусть условие выполнено на шаге k . Теперь воспользуемся динамической нумерацией вершин, которую мы задали при определении номера вершин в графе (см рис.2). На шаге $k+1$ у нас возможно две ситуации: второй индекс вершины равен 1 или не равен ей. Рассмотрим вариант, когда 1: на шагах k и меньше была задана часть КС, в ходе добавления которой было добавлено $j-1$ новых размерностей исходного метрического пространства R . Тогда задана какая-то прямая, на которой на расстоянии $(1+k/1000)$ находится предыдущая вершина, в случае 1 мы откладываем расстояние $(1+(k+1)/1000)$ от вершины k и получаем

МП той же размерности j , что и на предыдущем шаге, только с новой вершиной на расстоянии $(1+(k+1)/1000)$. Заметим что при таком добавлении в силу выбранной метрики новая добавленная вершина имеет для себя ближнюю - прошлую k -ю вершину, так как следующая от неё на той же прямой будет на расстоянии большем чем предыдущая. Теперь рассмотрим случай 2 и более: у нас есть пространство размерности J , тогда откладываем в объемлющем многомерном пространстве прямую, перпендикулярную исходному МП R^j , и на этой прямой в положительную полуось откладываем расстояние $(1+(k+1)/1000)$, что даёт там также расстояние до ближайшей вершины наименьшее в силу евклидовой метрики, а остальные условия не нарушаются.

Таким образом мы получаем построение МП и графа в этом МП, для которого исходный граф будет графом ближайших соседей.

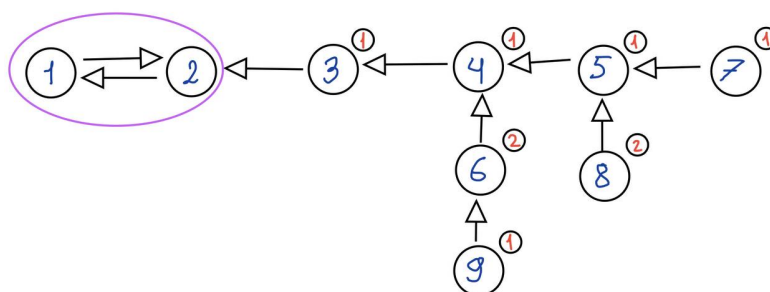


Рис. 2: Пример КС с нумерацией

Для иллюстрации построения такого графа, следует воспользоваться рис.3. На нём для выбранной компоненты связности получаем 3х мерное евклидово пространство. Алгоритм получения такого пространства описан в доказательстве теоремы. Отсюда получаем, что для любого графа

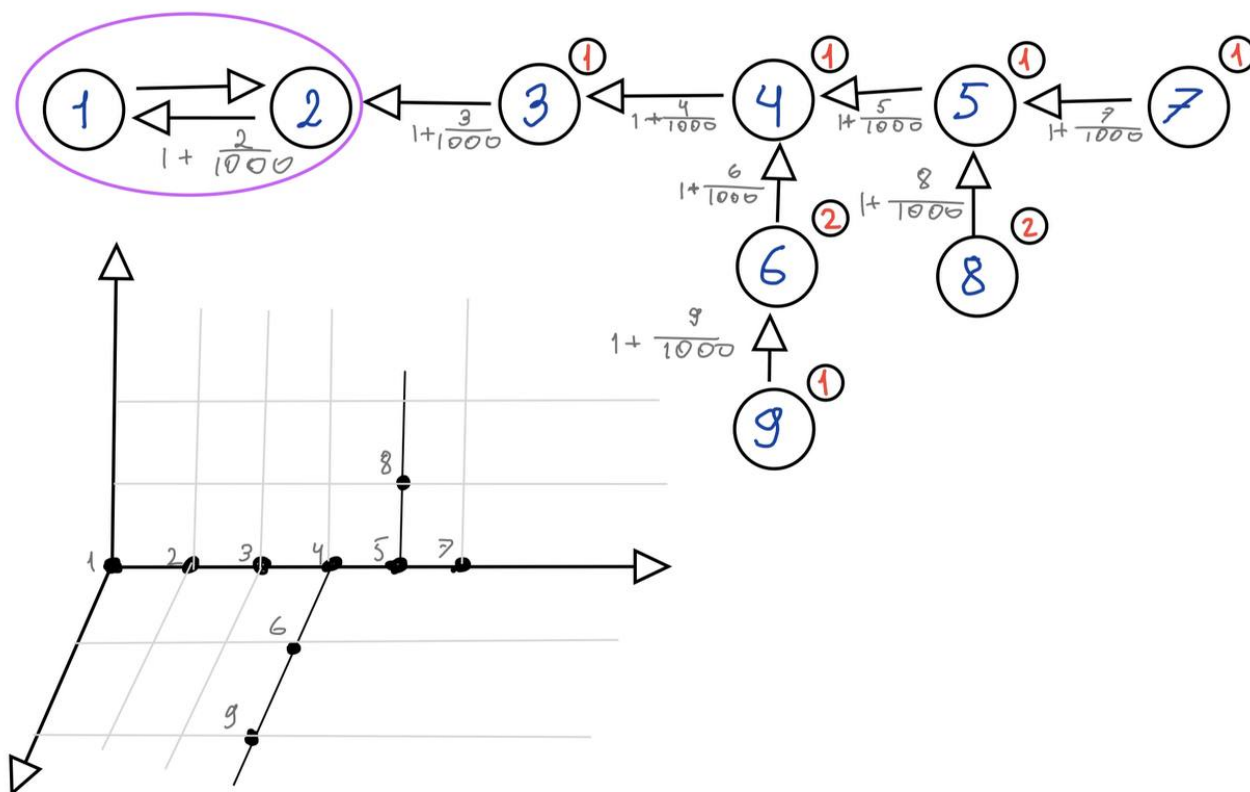


Рис. 3: Пример КС с нумерацией и расстоянием между вершинами

ближайших соседей существует МП, что в этом метрическом пространстве существует граф, для которого данный будет являться его графом ближайших соседей.

Теорема 3: Равномерная неуллучшаемость размерности пространства. Размерность польского пространства, определённого по методу теоремы 2 равномерно неуллучшаема числом.

Докажем эту теорему через пример. Граф, состоящий из цикла длины 2 и двух вершин, стрелки от которых приходят в одну из вершин цикла. Назовём эту вершину А, а приходящие вершины - В, С. Из теоремы 2 получаем размерность метрического пространства 2. Пусть существует одномерное метрическое пространство для которого существует граф, для которого рассматриваемый граф являлся бы ГБС. Тогда рассматриваем различные способы расположения точек В и С по отношению к А и приходим к выводу, что в силу неравенства треугольника существование МП размерности 1 невозможно, что и доказывает теорему о равномерной неуллучшаемости оценки из теоремы 2.

5 Граф вторых соседей

Проблема графа вторых соседей имеет обширные приложения и важна для изучения связей между вершинами графа. При рассмотрении графа вторых соседей, структура графа меняется и возникают новые структуры. Начнём рассмотрение со структуры графа вторых соседей. При изучении структуры ГБС в [2] было показано наличие цикла в каждой компоненте связности длины 2. При рассмотрении графа вторых соседей структура циклов в компонентах связности меняется. Для этого приведём в пример следующую теорему:

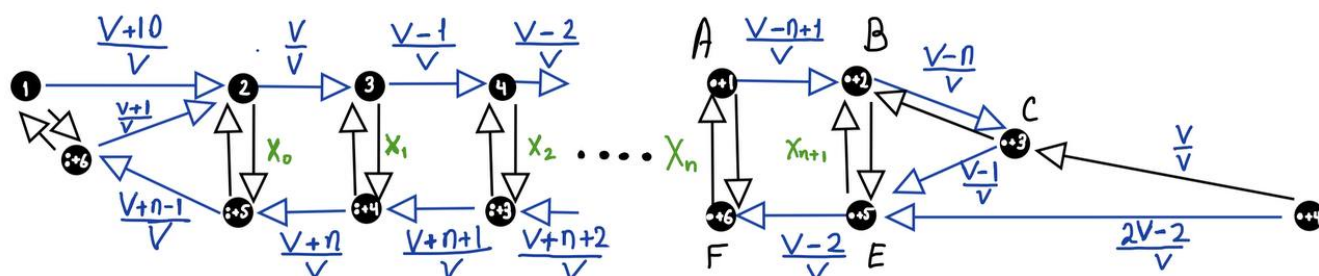
Предложение: Для большого чётного числа N вершин графа вторых соседей возможен цикл длины $N-2$.

Для доказательства приведём пример графа, для которого это выполнено. На рис.4 изображён граф с чётным числом вершин и связями между ними. Зададим расстояния между вершинами графа, где за N обозначим число вершин графа. Рассмотрение графа вторых соседей проводится для N большего или равного 3, так как рассмотрение для меньшего числа вершин не имеет смысла. Далее введём обозначения, требуемые для доказательства теоремы.

$$V = 100 \cdot N^2, \quad (5)$$

$$x = 1/2. \quad (6)$$

Чёрные стрелки отвечают за связи первых соседей, а синие - за связи вторых соседей. Для доказательства приведём расстояния на плоскости между вершинами графа. Расстояния между верхними вершинами будут равны $\frac{V+10}{V}, \frac{V}{V}, \frac{V-1}{V}, \frac{V-2}{V}$..., тогда как для нижнего будут значения: $\frac{V+n-1}{V}, \frac{V+n}{V}, \frac{V+n+1}{V}$.



и $2n + 5, 2n + 4, n + 6, n + 5, n + 4$. Вершины $2n + 6$ и $n + 3$ расположены вне этих прямых на заданном расстоянии от соседей. Расстояние между прямыми равно x , расстояния между парами $(2) - (2n + 5), (3) - (2n + 4)...$ задается расположением соответствующих вершин, начиная отсчёт с пары $(2) - (2n + 5)$, где отрезок, соединяющий эти две точки перпендикулярен заданным прямым. В силу заданных соотношений между вершинами получаем цикл, состоящий из синих стрелок длины $N-2$,

Замечание: монотонный поворот. Рассматривая на рисунке 4 переход синих стрелок от нижней прямой к верхней заметим монотонность по расстояниям между точками. Таким образом мы получили переход от убывающей последовательности расстояний к убывающей с сохранением последовательности убывания.

Замечание: ступенчатый поворот. Рассматривая на рисунке 4 переход синих стрелок от верхней прямой к нижней заметим отсутствие монотонности по расстояниям между точками. Таким образом мы получили переход от убывающей последовательности расстояний к убывающей без сохранения последовательности убывания.

Теорема 4: Для большого нечётного числа N вершин графа вторых соседей возможен цикл длины $N-1$.

Для доказательства рассмотрим граф, изображенный на рис.5.

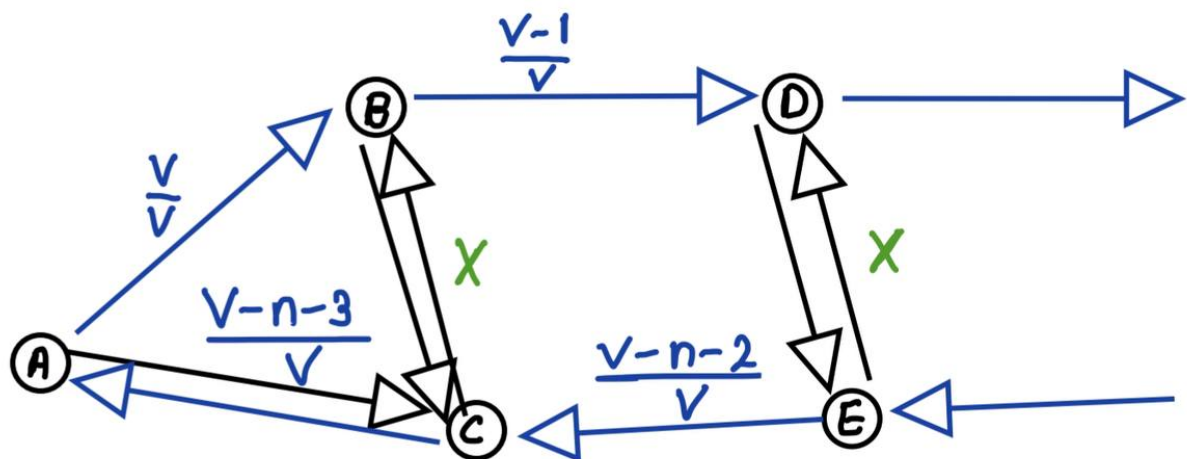


Рис. 5: Description

На этом графе чёрные стрелки отвечают за первых соседей, а синие за вторых. Расстояния между вершинами графа обозначены на рисунке. Расстояния x и V соответствуют расстояниям в Теореме 5. Так как расстояния

имеют возрастающую закономерность, и фиксированы x то черные стрелки будут поставлены именно так как на рисунке. Вершины $C, E \dots$ расположены на одной прямой, также как и верхний ряд вершин. По итогу получаем расстояния как на рис.5. Для дальнейшего доказательства стоит рассмотреть вершины: A, C, B и расположение синих стрелок для этих вершин. Для вершины C первым ближайшим будет B в силу того что x самый малый, а второй вершиной будет A , т.е. эта стрелка корректна. Тоже самое и для вершин A и B , что даёт "поворот" синих стрелок на левом конце цепочки. Правый конец цепочки имеет поворот, описанный в замечаниях к предложению. Откуда получаем наличие цикла длины $N-1$ для нечётного числа вершин.

Теорема 5: Для большого чётного числа N вершин графа вторых соседей возможен цикл длины не менее N .

Для доказательства этой теоремы воспользуемся Теоремой 4. Расставим вершины в такой же последовательности как и на рис.5, только для правого конца цепочки у нас будет ещё одна вершина, для которой сделаем расстояний как на рис.7. Аналогично теореме 4 получаем существование

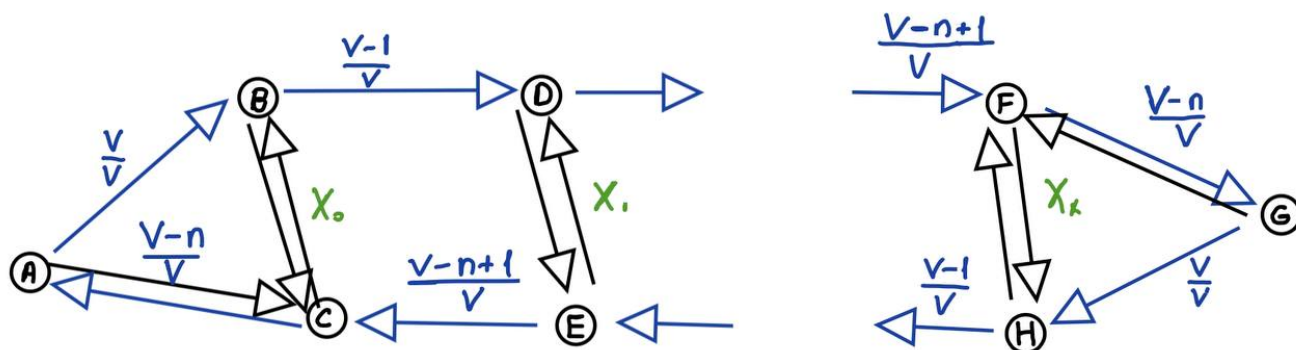


Рис. 6: Description

закругления на правом конце цепочки. Итого получили существование цикла длины N для чётного числа вершин.

6 Построение случайной матрицы и Гипотезы.

6.1 Построение случайной матрицы.

Как уже было упомянуто в 2.2, построение случайной матрицы возможно через определение расстояний между точками, брошенными в n -мерный единичный куб или через прямое задание элементов матрицы псевдослучайным образом. Для начала опишем способ построения через определения расстояний. В работах, связанных с поиском ближайших соседей (к примеру [6]) используется евклидова метрика для определения расстояний между точками. Будем использовать её. Точки будем кидать в n -мерный куб посредством проставления случайных координат для n -мерного вектора. Обозначим число брошенных точек за N . Так как в построении гипотезы важно число испытаний, то будем проводить испытания большое количество раз. Обозначим число испытаний за N_{test} . На каждом испытании мы получаем множество из N точек с n координатами, запишем матрицу $M_{test} \in M_{N \times N}$, где в ячейках матрицы будут евклидовы расстояния между точками в n -мерном кубе. Для построения графа ближайших соседей выделим для каждой точки ближайшую к ней. Для этого в каждом столбце матрицы M_{test} найдём минимальный элемент отличный от нуля, заменим его на 1, а все ненулевые оставшиеся элементы столбца обнулим. По итогу получим необязательно симметричную матрицу, на основе которой можно строить граф ближайших соседей. Процесс построения графа будет происходить по следующему принципу:

- 1 поиск необработанного столбца,
- 2 построение направленного ребра от номера найденного столбца к ненулевой вершине, указанной в этом столбце,
- 3 исключение столбца из очереди обработки.

После обработки матрицы, мы получили направленный граф, в котором вершинами являются брошенные точки, а рёбра направлены в сторону ближайшего для каждой из вершин соседа. Таким образом мы получили построение графа ближайших соседей, на основе случайной матрицы. Для этого графа нам интересно число компонент связности и число вершин в каждой из компонент связности.

Помимо этого способа построения матрицы, есть способ прямого задания элементов матрицы псевдослучайным образом.

Обратим внимание на способ построения графа ближайших соседей. Представленный ранее способ выкидывания точек в единичный куб и последующего определения расстояний описан в [7]. Принципиальным отличием является от прямого задания элементов матрицы является ожидаемое число компонент связности графа. Для графа, построенного на основе подсчёта расстояний это:

$$N_{comp} = \frac{3\pi}{8\pi + 3\sqrt{3}} \cdot N \approx 0.31 \cdot N, \quad (7)$$

в то время как для нашего способа построения это:

$$N_{comp} = 0.25 \cdot N. \quad (8)$$

6.2 Эксперимент.

Для получения зависимости числа компонент связности от числа элементов в компоненте связности проведём эксперимент. В ходе этого эксперимента будем строить матрицу одним (случайные элементы) из описанных способов, на основе этого графа будем получать серию из нескольких чисел: (N_{comp}, N_{elem}) , где N_{comp} - число компонент в элементе связности, а N_{elem} - число элементов в компоненте связности. Так как в графе возможно несколько компонент связности то за одну генерацию чисел псевдослучайным образом мы будем получать сразу несколько пар. Зададим матрицу размером $M \in M_{N/2 \times N}$, которая будет хранить данные о числе пар. При получении новой пары (N_i, N_j) будет увеличивать на один элемент m_{ij} матрицы М. После проведения эксперимента получим матрицу, которую можно визуализировать в трёхмерном пространстве, где на плоскости XY будет число компонент связности и число элементов в каждом из компонент связности, а по оси Z будет число пар для каждого из значений. Данная визуализация изображена на рис.7:

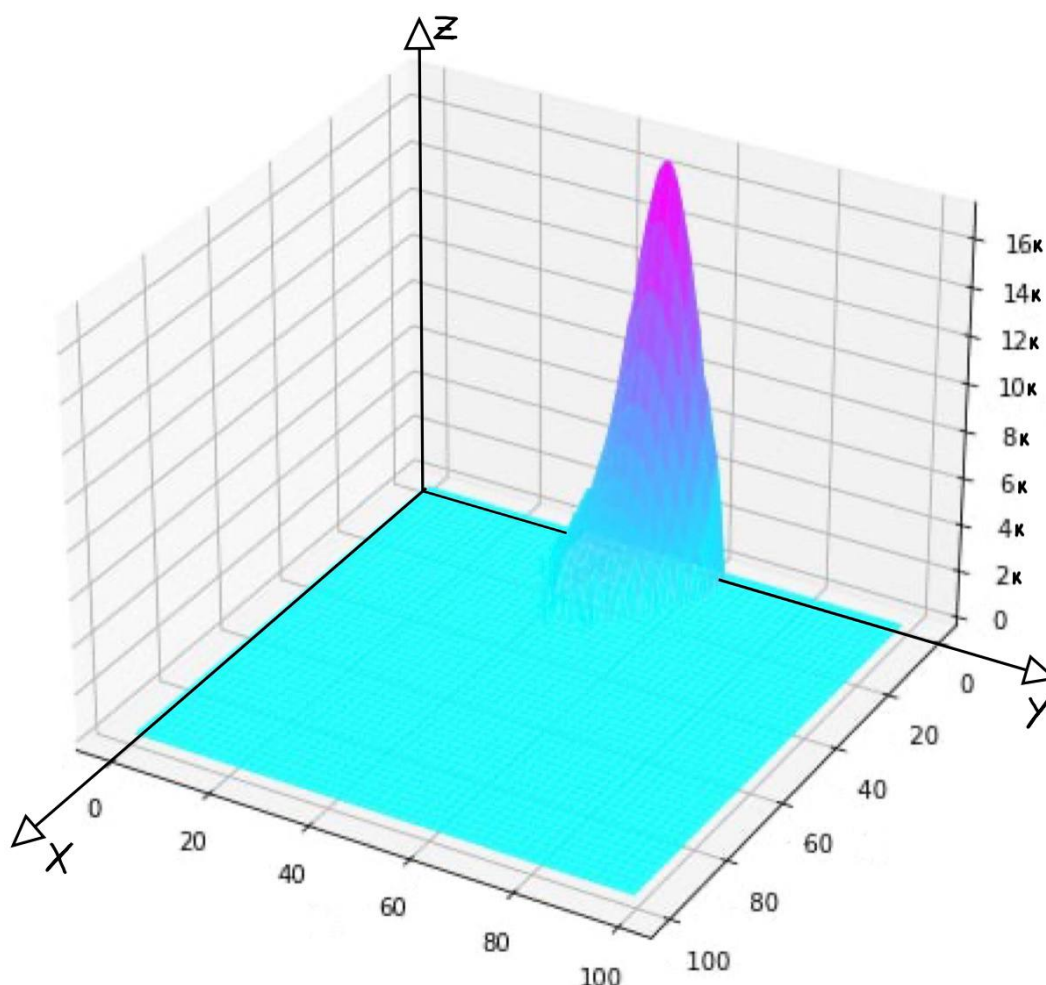


Рис. 7: Description

Эксперимент на рис.7 был проведён для 200 точек, брошенных в 2х мер-

ный единичный куб при общем числе испытаний в 100 тысяч штук. Нам интересно сечение этого графика различными линиями уровня. Для определения этих сечений нанесём на плоскость координаты точек матрицы M , по следующему правилу:

$$M_{layer} = \{(i, j) | z_1 \leq m_{ij} \leq z_2, \forall m_{ij} \in M\}, \quad (9)$$

где z_1 и z_2 задают границы уровня на рис.7. Эти точки образуют следующее сечение:

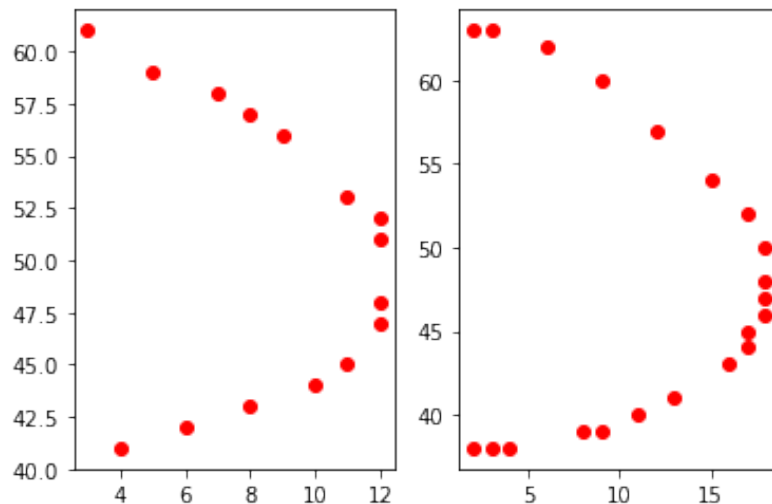


Рис. 8: Сечение между 5500 и 6000 значений по оси слева и между 2700 и 3000 справа.

6.3 Гипотезы.

Согласно полукруговому закону Вигнера плотность числа собственных значений на интервале подчиняется закон полукруга [8]. Так как в рассмотрении участвует симметричная вещественная положительная матрица, то логично предположить, что какие-то свойства графа ближайших соседей, построенного на основе этой случайной матрицы подчиняются полукруговому закону Вигнера. Основной гипотезой, выдвигаемой на основе экспериментальных и теоретических данных является:

Гипотеза. Сечерние графика, построенного на основе эксперимента*, представляет собой полукруг в пределе количества точек.

Рассмотрим правдоподобность этой гипотезы. Для этого стоит рассмотреть динамику изменения сечерния в зависимости от числа точек, участвующих в расчёте. Для этого выполним те же расчёты уже для 90, 120, 150, 200, 300 и 500 точек, участвующих в расчётах. Результаты сечейний представлены ниже:

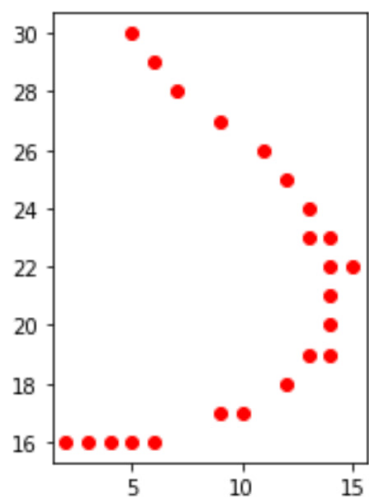


Рис. 9: 90 точек.

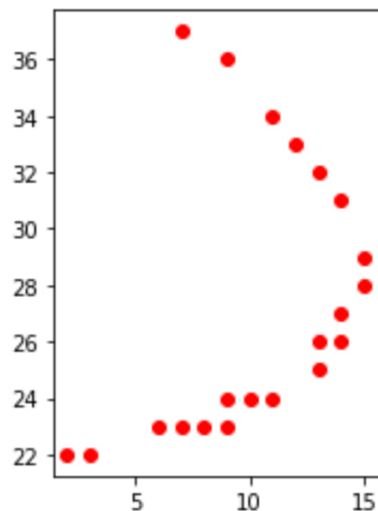


Рис. 10: 120 точек.

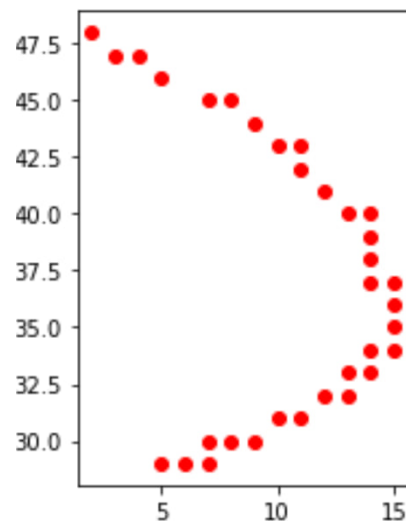


Рис. 11: 150 точек.

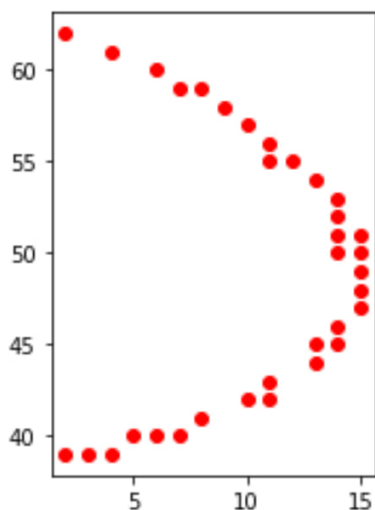


Рис. 12: 200 точек.

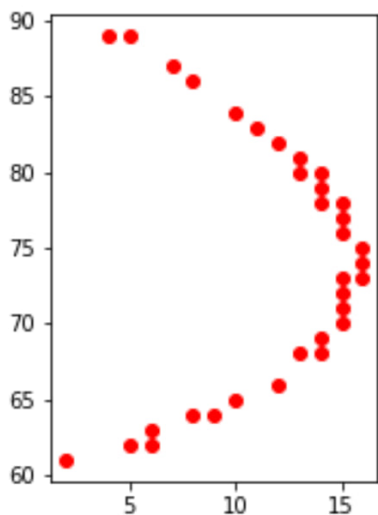


Рис. 13: 300 точек.

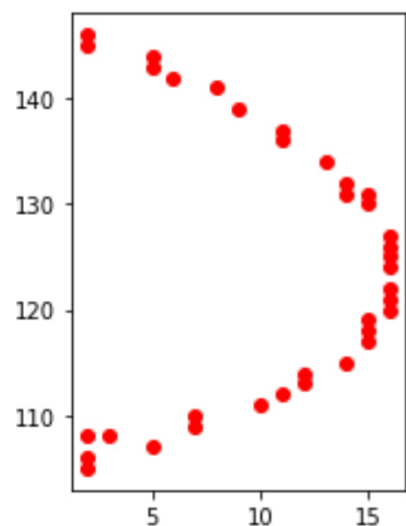


Рис. 14: 500 точек.

На основе рис.9-рис.14 можно сделать вывод о корректности постановки гипотезы.

7 Заключение

В работы мы рассмотрели случайные матрицы, случайные матрицы расстояний, графы ближайших соседей и способы их построения. Для графа, обладающего заданными свойствами графа ближайших соседей мы получили существование польского пространства, реализующего его. Причем получили равномерную ограниченность на размерность пространства, реализующего данный граф. Помимо этого мы рассмотрели граф k -ближайших соседей, обладающий рядом полезных свойств, которые нам удалось доказать. Основным свойством является существование цикла длины N для графа вторых соседей, построенного на N точках.

Помимо этого мы определили принципиальную разницу между способом построения графа ближайших соседей. Показали что построения на основе подсчёта расстояний имеют матожидание числа компонент связности равное:

$$N_{comp} = \frac{3\pi}{8\pi + 3\sqrt{3}} \cdot N \approx 0.31 \cdot N, \quad (10)$$

в то время как для построения графа, при задании расстояний случайным образом первым моментом будет являться $N_{comp} = 0.25 \cdot N$, где за N обозначено количество вершин графа.

В экспериментальной части работы мы познакомились со способом задания эксперимента посредством генерации псевдослучайных чисел на языке python с последующей обработкой полученных данных, подсчётом моментов и сечений соответствующих графиков. Нам удалось сформулировать гипотезу, которая утверждает стремление сечения графика к эллиптической форме.

8 Литература

- 1 А.М. Вершик, Случайные метрические пространства и универсальность.
- 2 Препринт Орлова
- 3 А.А. Кислицын, М.В. Гогугев Вычислительная экспансия в графы ближайших соседей: статистики и размерности пространства
- 4 А.А Кислицын. Ю.Н. Орлов. Исследование статистик графов ближайших соседей // Препринты ИПМ им. М.В. Келдыша. 2021. № 85. 23 с.
- 5 А.А. Кислицын Исследование статистик графов ближайших соседей // Математическое моделирование, 2022. Т. 34. № . 8. С. 110-126.
- 6 L. Prokhorenkova A. Shekhovtsov, 2021. Graph-based Nearest Neighbor Search: From Practice to Theory
- 7 D.Eppstein M.S. Paterson F.F.Yao, 2000. On Nearest-Neighbor Graphs
- 8 Wigner E. On the distribution of the roots of certain symmetric matrices. // Ann. of Math., 67 (1958), 325-328.
- 9 А.С. Кристаль [https : //github.com/st2257st2257/bachelors_degree](https://github.com/st2257st2257/bachelors_degree)