

О СПЕКТРАЛЬНЫХ ПОРТРЕТАХ МАТРИЦ ИНЦИДЕНТНОСТИ ГРАФОВ БЛИЖАЙШИХ СОСЕДЕЙ¹⁾

© 2024 г. А. А. Кислицын^{1,*}¹125047 Москва, Миусская пл., 4, ИПМ им. Келдыша РАН, Россия

*e-mail: alexey.kislitsyn@gmail.com

Поступила в редакцию 02.04.2024 г.

Переработанный вариант 02.04.2024 г.

Принята к публикации 05.05.2024 г.

В работе исследуется возможность применения метода С.К. Годунова построения спектрального портрета матрицы для оценки ранга матриц специального вида, которые появляются в ряде приложений, таких, как анализ структур графов ближайших соседей, теория конечных автоматов и оценка спектра разреженных матриц. Описан вычислительный алгоритм генерации ансамбля матриц случайных расстояний и связанных с ним графов ближайших соседей. На основе вычислительного эксперимента оценены параметры распределения степеней вершин случайных графов ближайших соседей. Оценки удается получить вследствие того, что указанное распределение не зависит от функции распределения случайных расстояний и является многомерным нормальным распределением. Доказано, что ранг матрицы инцидентности графа ближайших соседей равен суммарному числу вершин со степенями 0 и 1 по входящим ребрам, и получено распределение ранга такой матрицы. Показано, что в данной задаче определения ранга матрицы весьма эффективным является также метод, основанный на анализе распределения степеней вершин. Библ. 17. Фиг. 4.

Ключевые слова: граф ближайших соседей, распределение степеней вершин, спектральный портрет, распределение ранга матрицы.

DOI: 10.31857/S0044466924080189, EDN: XZTUET

1. ВВЕДЕНИЕ

Во многих задачах вычислительной линейной алгебры необходимо иметь оценки того, в каких пределах лежит возмущение спектра матрицы при возмущении ее элементов. Это связано с тем, что на практике при построении разностных схем для решения уравнений математической физики появляются матрицы коэффициентов, характеризующих свойства системы в вычислительной ячейке. Ими могут быть, например, коэффициенты теплопроводности или вязкости, получаемые в ходе физического эксперимента, и потому известные с некоторой ошибкой. В результате может оказаться, что свойства численного решения неожиданно сильно зависят от точности входных данных, которые можно трактовать как параметры со случайным возмущением. Возникающие при этом проблемы были описаны в известных монографиях [1, 2].

В дальнейшем выяснилось, что исследование устойчивости спектра матрицы при случайном возмущении ее элементов требуется проводить также в ряде статистических задач для оценки достаточности длины выборки, корректности модели распознавания образов, оценки ранга случайной матрицы. Примеры такого использования спектральных портретов стохастических матриц содержатся в книге [3].

Современное состояние исследований представлено в монографии С.К. Годунова [4], которой мы и будем следовать в части основных определений и методологии.

Представляет интерес использование спектральных портретов матриц, возникающих при анализе случайных разреженных матриц больших размеров. Такие матрицы возникают, в частности, как матрицы инцидентности графов ближайших соседей при моделировании структурных связей тех или иных множеств.

Анализ графов ближайших соседей (далее ГБС) представляет собой один из широко используемых методов кластеризации данных [5]. Графом *ближайших соседей* называется ориентированный граф, определенный для множества E точек в метрическом пространстве. Каждая точка данного множества является вершиной графа. Направленное ребро из точки A в точку B проводится для той точки B множества, расстояние до которой от

¹⁾ Работа выполнена при финансовой поддержке РНФ (код проекта 23-71-10055).

точки A минимально. Обозначим через $\rho_A(X)$ расстояние от данной точки $A \in E$ до некоторой другой точки $X \in E$. Тогда точка B , к которой проводится ребро из точки A , определяется условием

$$B = \arg \min_{\substack{X \in E \\ X \neq A}} \rho_A(X).$$

В большинстве случаев целью анализа является установление структурных связей между объектами, которым были поставлены в соответствие наборы чисел и введена мера близости в некотором подходящем пространстве. Тогда практическую важность приобретает вопрос о вариабельности установленных эмпирических закономерностей, поскольку разные выборки однотипных объектов будут порождать, вообще говоря, разные графы связей. В этом смысле граф связей является случайным, поскольку случаен сам отбор объектов в анализируемую группу.

Теория случайных графов развита очень подробно в части случайных графов, которые обычно трактуются в смысле Эрдёша–Реньи [6], когда вершины соединяются между собой по заданному вероятностному закону. Такие графы активно изучались в контексте решения многих прикладных задач (см., например, [7–9]). Асимптотические свойства случайных графов также анализировались в рамках конкретных вероятностных моделей применительно к различным онтологиям, социальным и транспортным сетям и аналогичным объектам (см. работы [10–13]). Однако в области статистического анализа графов конструктивных моделей и результатов существенно меньше. Это связано с тем, что выборка графов сама по себе не имеет места, поскольку граф визуализирует свойство изучаемой группы объектов. Поэтому понятие выборки адресуется именно группе объектов, которые должны быть однородными по изучаемому свойству в статистическом смысле.

Основная проблема состоит в том, что анализ одного конкретного графа не позволяет оценить вероятность его реализации как элемента из определенного множества случайных графов, поскольку нужное множество не формализовано. Исключением является граф ближайших соседей, который строится на основе матрицы попарных расстояний между точками. В этом случае функция распределения расстояний известна, что позволяет выяснить, какие графы ей отвечают. В работах [14–16] были получены результаты о статистике структур графов ближайших соседей. В основе вычислительного метода лежит теорема о независимости распределения этих структур от функции распределения расстояний между вершинами графа. Этот факт позволил получить в результате численных экспериментов табличные значения вероятностей, характеризующих ансамбль ГБС: распределение графов по числу несвязных фрагментов, фрагментов по числу вершин и вершин по степеням входящих ребер. Используя эти результаты, можно построить распределение ранга матрицы инцидентности и получить новый критерий независимости выборочных данных.

2. СПЕКТРАЛЬНЫЙ ПОРТРЕТ МАТРИЦЫ

Приведем здесь определения основных понятий, требующихся далее, следуя монографии С.К. Годунова [4].

Число λ принадлежит ε -спектру $\Lambda_\varepsilon(A)$ матрицы A , если существует такая возмущающая ее матрица Δ , что $\|\Delta\| \leq \varepsilon \|A\|$ и $\det(\lambda I - A - \Delta) = 0$, где I — единичная матрица.

Сингулярными числами σ_i матрицы A называются собственные значения эрмитовой матрицы $A^+ A$, где знак плюс сверху означает эрмитово сопряжение. Все эти числа неотрицательны. Занумеруем их в порядке возрастания, так что для матрицы порядка N минимальное сингулярное число есть σ_1 , а максимальное σ_N .

Числом обусловленности невырожденной матрицы A называется величина

$$\mu(A) = \frac{\sigma_N(A)}{\sigma_1(A)} = \|A\| \cdot \|A^{-1}\|. \quad (2.1)$$

Имеет место оценка для числа обусловленности возмущенной матрицы:

$$|\mu(A + \Delta) - \mu(A)| \leq \frac{2\mu(A)}{1 - \mu(A) \frac{\|\Delta\|}{\|A\|}} \cdot \frac{\|\Delta\|}{\|A\|}. \quad (2.2)$$

Эта оценка означает, что при малом возмущении, когда $\|\Delta\| \ll \|A\|/\mu(A)$, число обусловленности возмущенной матрицы также меняется мало.

Резольвентой матрицы A называется матрица

$$R(\lambda) = (\lambda I - A)^{-1}. \quad (2.3)$$

В терминах резольвенты ε -спектр определяется следующим образом: число λ принадлежит ε -спектру $\Lambda_\varepsilon(A)$ матрицы A , если

$$\|R(\lambda)\| \geq \frac{1}{\varepsilon \|A\|}. \quad (2.4)$$

На этой формуле основан практически применяемый численный алгоритм определения областей в комплексной плоскости параметра λ , отвечающих спектральным пятнам в зависимости от величины возмущения ε .

При исследовании расположения точек спектра представляют интерес замкнутые гладкие кривые γ_ε , представляющие изолинии ε -спектра. Контур γ_ε разбивает весь ε -спектр $\Lambda_\varepsilon(A)$ на две части, лежащие внутри и вне его. В этом смысле контур γ_ε осуществляет дихотомию ε -спектра матрицы. Качество дихотомии $\kappa_\gamma(A)$ оценивается нормой квадрата резольвенты (2.3) на данной кривой:

$$\kappa_\gamma(A) = \frac{\|A\|^2}{l_\gamma} \oint_\gamma \|R(\lambda)\|^2 d\lambda. \quad (2.5)$$

Здесь l_γ есть длина контура γ . Величина $\kappa_\gamma(A)$ выбрана как индикатор точности разделения спектра потому, что если на некоторой кривой γ нет точек спектра $\lambda(A)$, то норма резольвенты на такой кривой конечна: $\|R(\lambda)\|_\gamma < \infty$, как и интеграл от нее по этой кривой.

Если внутри области, ограниченной кривой γ_ε , оказалось несколько собственных значений, то с указанной точностью ε их естественно считать совпадающими. Тогда подпространство с базисом из собственного и присоединенных векторов для такого кратного собственного значения будет инвариантным подпространством для оператора A . Проектор Π на это инвариантное подпространство определяется формулой

$$\Pi_\gamma = \frac{1}{2\pi i} \oint_\gamma R(\lambda) d\lambda. \quad (2.6)$$

В частности, если контур γ охватывает все собственные значения, то всякая аналитическая функция от матрицы, будучи определенной на ее спектре, имеет интегральное представление

$$f(A) = \frac{1}{2\pi i} \oint_\gamma f(\lambda) R(\lambda) d\lambda. \quad (2.7)$$

Удобно рассматривать радиальную дихотомию, т.е. дихотомию, задаваемую кривой $\lambda = re^{i\varphi}$ при фиксированном значении r . Тогда параметр дихотомии $\kappa_r(A)$ является нормой эрмитовой матрицы $H_r(A)$, имеющей интегральное представление

$$H_r(A) = \frac{1}{2\pi} \int_0^{2\pi} (A^+ - re^{-i\varphi} I)^{-1} (A - re^{i\varphi} I)^{-1} d\varphi, \quad \kappa_r(A) = \|A\|^2 \cdot \|H_r(A)\|. \quad (2.8)$$

Интеграл в (2.8) сходится только в том случае, если на окружности $\lambda = re^{i\varphi}$ нет собственных значений матрицы A . Это представление используется при численном нахождении ε -спектра матрицы.

3. ОСНОВНЫЕ СВОЙСТВА ГРАФОВ БЛИЖАЙШИХ СОСЕДЕЙ

Мы будем предполагать единственность ближайшего соседа для каждой точки множества. Тогда из определения ГБС следуют его основные структурные свойства: с каждой вершиной связано ровно одно выходящее ребро, число вершин графа равно числу ребер, каждый связный ГБС содержит ровно один цикл. Для дальнейшего существенны следующие утверждения.

Пусть $n_p(N)$ есть число вершин степени p в графе ближайших соседей из N вершин. Тогда

$$\sum_{p=0}^{N-1} n_p(N) = N, \quad \sum_{p=0}^{N-1} pn_p(N) = N. \quad (3.1)$$

Отсюда следует, что число вершин нулевой степени равно

$$n_0 = \sum_{p=2}^{N-1} (p-1)n_p, \quad (3.2)$$

а сумма числа вершин нулевой и первой степени равна

$$n_0 + n_1 = N - \sum_{p=2}^{N-1} n_p. \quad (3.3)$$

Справедливы также следующие утверждения.

Теорема 1 (см. [14]). *Вероятность реализации структуры ГБС не зависит от распределения расстояний между точками.*

Утверждение следует из того, что монотонное преобразование одного стационарного распределения вероятностей в другое не меняет упорядоченности между элементами выборки, т.е. не влияет на структуру ГБС.

Теорема 2 (см. [16]). *Статистика результата генерации ансамбля графов ближайших соседей не зависит от выполнения неравенства треугольника для элементов матрицы, которые могут трактоваться как расстояния, а могут — как независимые случайные числа.*

Доказательство основано на том, что предъявляется алгоритм построения матрицы именно расстояний в некотором метрическом пространстве, после чего, используя теорему 1, делается преобразование получающегося распределения расстояний к равномерному распределению, при котором упорядоченность между элементами матрицы сохраняется, но неравенство треугольника уже не выполняется.

Новый результат, обсуждению которого посвящена настоящая работа, состоит в следующем.

Теорема 3. *Сумма числа нулевых и первых степеней вершин графа ближайших соседей равна рангу соответствующей матрицы инцидентности.*

Доказательство. Рассмотрим ГБС, порожденный некоторой матрицей расстояний $N \times N$. Сопоставим этой матрице расстояний матрицу инцидентности для соседа порядка k . Эта матрица того же порядка $N \times N$, что и исходная матрица случайных расстояний, в которой все элементы строки i , кроме одного, отвечающего соседству вершины в строке i и вершины в столбце j , заменены нулями. Отмеченный же элемент 1_{ij} положен равным единице. Поскольку в силу единственности соседа данного порядка в каждой строке матрицы есть ровно один ненулевой элемент и сумма всех таких элементов равна порядку матрицы N , то необходимым и достаточным условием невырожденности матрицы является в данном случае отсутствие нулевого столбца. Необходимость условия очевидна (для невырожденности необходимо отсутствие нулевого столбца), а достаточность следует из разложения определителя по i -й строке, что сводит вычисление определителя матрицы $N \times N$ к вычислению определителя матрицы $(N-1) \times (N-1)$ с точностью до знака $(-1)^{i+j}$. Совпадающих ненулевых столбцов в матрице инцидентности нет, так как иначе это означало бы, что у некоторой точки более одного ближайшего соседа. Следовательно, ранг матрицы инцидентности равен ее размеру за вычетом числа нулевых столбцов. Нулевые столбцы могут появиться только в том случае, если в каких-то других столбцах число единиц больше одной. Наличие, например двух единиц в столбце j означает, что вершина j является ближайшей для некоторых двух вершин. Следовательно, нулевых столбцов в матрице инцидентности столько, сколько имеется вершин степени 2, 3, и т.д. Поскольку же $\sum_{p=0}^{N-1} n_p = N$, то число ненулевых столбцов равно $N - \sum_{p=2}^{N-1} n_p = n_0 + n_1$. Это и есть порядок максимального отличного от нуля минора. Теорема 3 доказана.

Следствием этого утверждения является то, что распределение ранга матрицы инцидентности равно распределению суммы двух случайных величин — числа вершин со степенями 0 и 1 по входящим ребрам.

4. АЛГОРИТМ ГЕНЕРАЦИИ МАТРИЦ СЛУЧАЙНЫХ РАССТОЯНИЙ

Для численного анализа статистик структур ГБС был разработан алгоритм, основанный на теоремах 1 и 2. При сборе статистик графов первых ближайших соседей были использованы следующие особенности именно ГБС структур.

Поскольку ближайший сосед ровно один, то вместо матрицы смежности достаточно хранить всего одну строку, в k -й ячейке которой записан номер $j(k)$ вершины j , которая является ближайшей для вершины k .

Пересчет числа фрагментов начинается с первой ячейки. В первый фрагмент попадают вершины, связанные цепочкой: $1 \rightarrow j_1(1) \rightarrow j_2(j_1(1)) \rightarrow j_3(j_2(j_1(1))) \rightarrow \dots$. Цепочка продолжается до тех пор, пока в ней не появляется номер вершины, уже отмеченный ранее. После этого найденная последовательность номеров вершин записывается как фрагмент 1. Эти номера вычеркиваются из общей строки. Затем начинается аналогичный процесс для ячейки, номер которой следует за 1 (с учетом вычеркивания). Если на каком-то шаге появляется номер $j(k)$, найденный ранее, то вся эта цепочка присоединяется к соответствующему фрагменту. В результате собираются данные о количестве вершин в каждом фрагменте, что позволяет построить по достаточно большому числу таких экспериментов распределение фрагментов по числу вершин.

Распределение вершин по степеням как для отдельного фрагмента, так и для графа в целом строится по частотам встречаемости номеров вершин, являющихся чьими-то соседями: это количество номеров, встретившихся в строке $j(k)$ 0 раз, 1 раз, 2 раза и т.д. В итоге после реализации K экспериментов по генерации матрицы $N \times N$ случайных равномерно распределенных чисел получаем набор $(n_0^1, \dots, n_0^K)$ вариантов для числа вершин с нулевой степенью, что позволяет построить эмпирическое распределение вероятности количества

таких вершин $f_0(n)$, так что $\sum_{n=0}^N f_0(n) = 1$. Из тех же экспериментов получаем набор $(n_1^1, \dots, n_1^K)$ вариантов для числа вершин с первой степенью и соответствующее распределение $f_1(n)$ и т.д.

С учетом однотипности производимых вычислительных процедур целесообразно

перейти к параллельной обработке данных. Алгоритм для параллельной реализации выглядит следующим образом. На первом шаге задаются инициализирующие параметры. Например: количество реализаций, размер матрицы графа, количество потоков. Далее происходит настройка главного процесса, который будет раздавать задачи всем остальным. Инициализируется набор процессов, на которых непосредственно будут происходить вычисления, и на них запускается реализация основного вычислительного блока. В основном вычислительном блоке происходят следующие операции:

- генерируется стохастическая симметрическая матрица заданных размеров;
- в каждом столбце проводится упорядочение значений по близости к первому элементу столбца, что позволяет для каждого первого элемента столбца получить вектор его соседей, отсортированный по удаленности;
- формируется вектор соседей заданной удаленности;
- по сформированным векторам соседей выстраивается ориентированный граф с использованием библиотеки networkx;
- собирается статистика по количеству связанных компонент и количеству вершин;
- рассчитывается распределение вершин графа по степеням.

На последнем шаге проводится проверка контрольных сумм, общего числа сгенерированных вершин и количества графов.

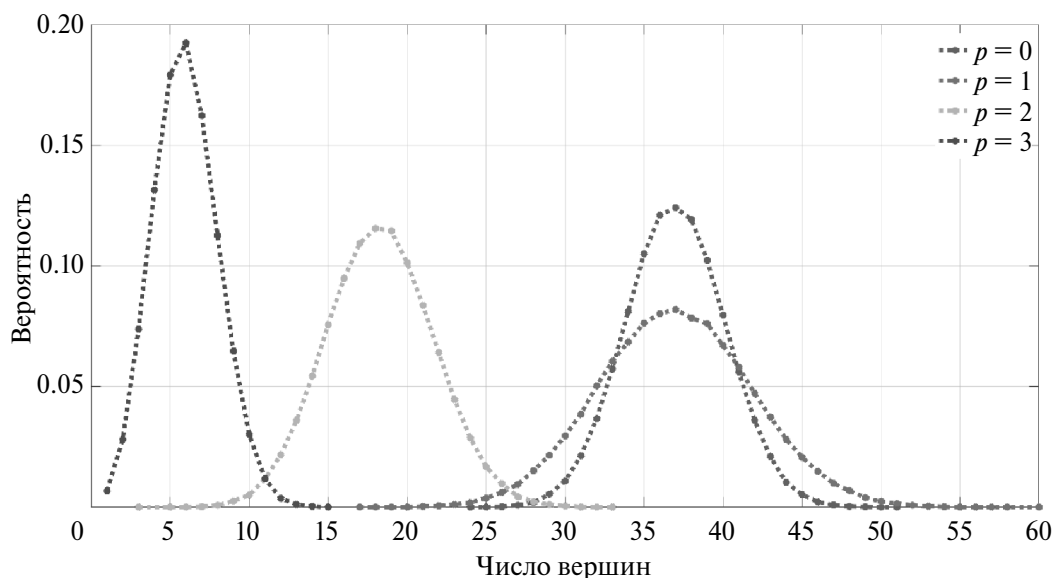
Подчеркнем, что в силу теоремы 2 нет необходимости генерировать расстояния между случайными точками в евклидовом n -мерном пространстве, можно сгенерировать непосредственно матрицу случайных чисел. Это заметно сокращает вычисления, а также позволяет проводить анализ ближайших соседей в различных квазинормах, таких, например, как расстояние Кульбака–Лэйблера.

5. РАСПРЕДЕЛЕНИЕ РАНГА МАТРИЦЫ ИНЦИДЕНТНОСТИ ГРАФОВ БЛИЖАЙШИХ СОСЕДЕЙ

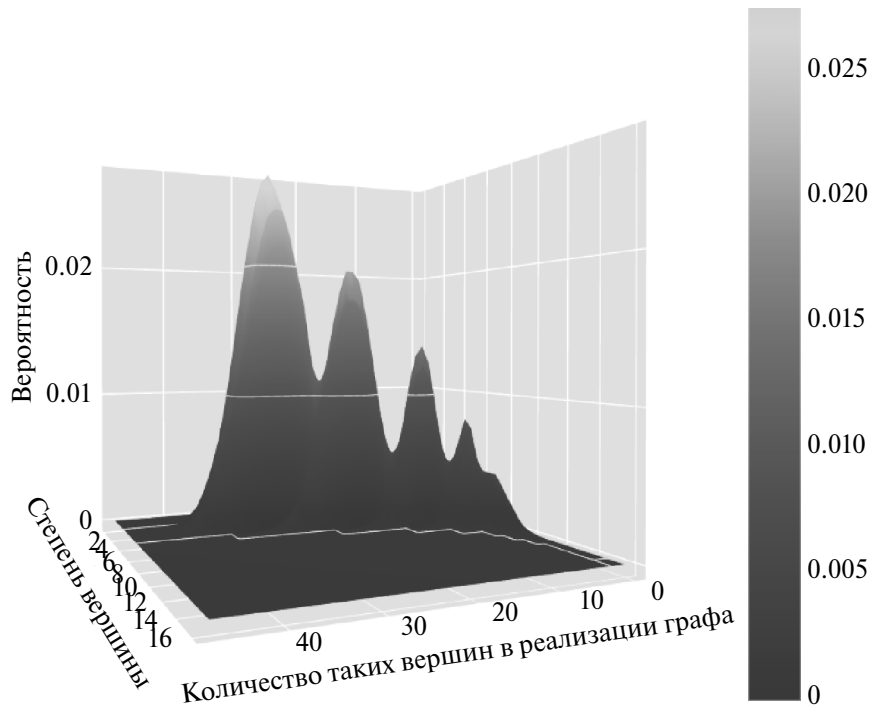
Важная информация о структуре графов ближайших соседей содержится в распределении вершин по степеням. В результате численного моделирования выяснилось, что зависимость этого распределения от числа вершин обладает определенным самоподобием, что позволяет численно найти характеризующие его константы.

На фиг. 1 приведено распределение вершин ближайших соседей по степеням для графа с числом вершин $N = 100$ по результатам 1 млн экспериментов. Показаны вероятности $f_P(n)$ того, что число вершин со степенью P равно n .

Если расположить эти распределения на одной плоскости в виде поверхности $F(p, n)$, которая есть совокупность распределений $f_p(n)$, то получим фазовое пространство возможных реализаций графа. При этом воз-



Фиг. 1. Распределение вершин по степеням для $N = 100$.



Фиг. 2. Распределение числа вершин графа по степеням для $N = 100$.

можно не любая траектория $(p, n(p))$, а лишь такая, для которой выполняются условия (3.1). Соответствующая поверхность $F(p, n)$ приведена на фиг. 2 для $N = 100$.

Совместное распределение степеней вершин ГБС асимптотически нормально. Это следует из того, что при случайном размещении n частиц по N ячейкам вероятность $P(\xi_1 = k_1, \dots, \xi_N = k_N | \xi_1 + \dots + \xi_N = n)$ совместного распределения заполнения ячеек есть полиномиальное распределение схемы размещения (n, N) :

$$P(\xi_1 = k_1, \dots, \xi_N = k_N | \xi_1 + \dots + \xi_N = n) = \frac{n!}{N^n k_1! \dots k_N!}, \quad (5.1)$$

которое при больших N стремится к многомерному нормальному распределению. Важно подчеркнуть, что теоретически параметры этого распределения ранее определены не были. По итогам вычислительного эксперимента выяснилось, что эмпирическое совместное распределение числа вершин по степеням $\{0, 1, \dots, p\}$ может быть приближенно описано распределением вида

$$f(n_0, n_1, \dots, n_p) = \frac{1}{\sqrt{\det C}} \left(\frac{N}{\pi} \right)^{(p+1)/2} \exp \left(-N \sum_{i,j=0}^p C_{ij}^{-1} (n_i/N - \mu_i)(n_j/N - \mu_j) \right), \quad (5.2)$$

где $\det C$ есть определитель ковариационной матрицы, C_{ij}^{-1} — элементы обратной ковариационной матрицы, n_j — число вершин со степенью j , μ_j — средняя доля вершин, имеющих степень j . Поскольку показатель экспоненты имеет порядок N , то в силу быстрого стремления плотности распределения к нулю при малых значениях n_j и при значениях порядка N приближение нормировки по всему p -мерному пространству от $-\infty$ до $+\infty$ по каждой координате допустимо, хотя формально аргументы распределения в (5.2) неотрицательны.

По результатам 1 млн генераций графов были получены следующие значения μ_j , не зависящие от N :

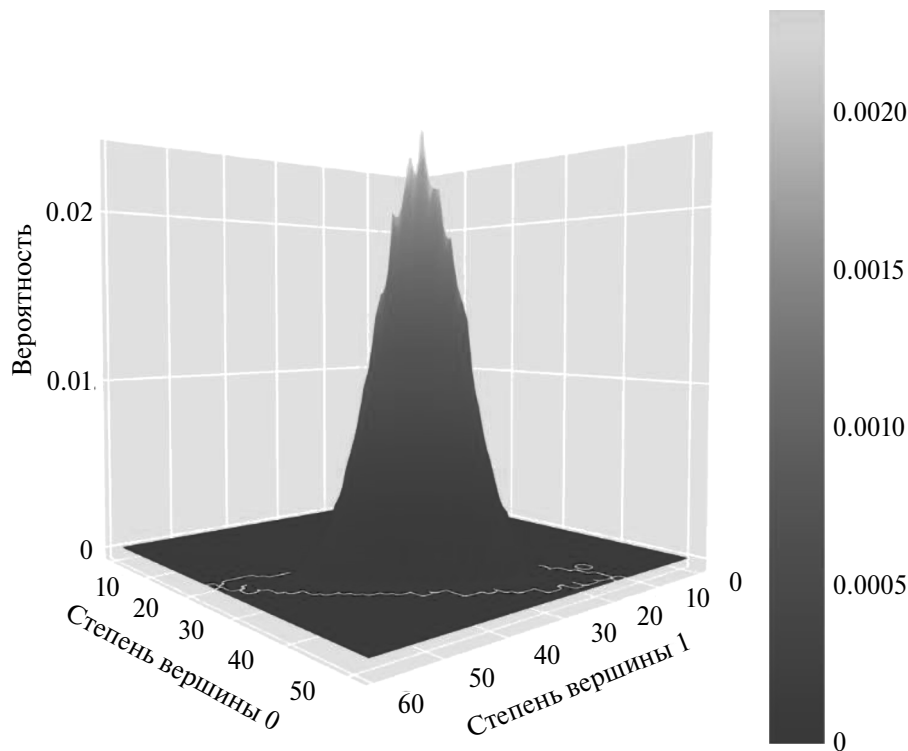
$$\mu_0 = 0.367; \quad \mu_1 = 0.367; \quad \mu_2 = 0.184; \quad \mu_3 = 0.061; \quad \mu_4 = 0.015; \quad \mu_5 = 0.003; \quad \mu_6 = 0.0005. \quad (5.3)$$

Статистическая достоверность предъявленных результатов выше 0.99.

Из совместного распределения (5.2) можно получить практические аппроксимации для плотностей распределения числа вершин по степеням. Например, для нулевой, первой и второй степеней в графе из N вершин имеем формулы:

$$f_p(n) = \sqrt{\frac{A_p}{\pi N}} \exp \left(-\frac{A_p}{N} (n - \mu_p N)^2 \right), \quad (5.4)$$

$$A_0 = 4.9, \mu_0 = 0.37; \quad A_1 = 2.1, \mu_1 = 0.37; \quad A_2 = 4.5, \mu_2 = 0.18.$$



Фиг. 3. Совместное распределение числа вершин со степенями 0 и 1 для $N = 100$.

В силу условия (3.1) распределения числа вершин по степеням коррелированы. Совместное распределение числа вершин степеней 0 и 1 для $N = 100$ показано на фиг. 3.

Этот важный аспект находит свое приложение в анализе матрицы инцидентности. Такие матрицы рассматриваются не только в теории графов, но и в теории конечных автоматов (см. [17]). Обычно сами матрицы для описания конечных автоматов постоянны, а случайным является подаваемый на них сигнал, так что по анализу выходного сигнала делается попытка определения указанной матрицы. В нашей задаче в некотором смысле обратная ситуация — случайной является матрица. Также распределение числа вершин с нулевой и первой степенями по входящим ребрам имеет отношение к спектральной теории случайных графов.

Несмотря на конструктивную простоту доказательства теоремы 3, в качестве следствия из нее ниже получен результат, который следует из (5.4). Приближенными в этой формуле являются коэффициенты A_p и μ_p , найденные в результате численного моделирования. Сама же формула нормального распределения является асимптотически точной по N для распределения степеней вершин.

Носитель совместного распределения для двух случайных величин (числа вершин со степенями 0 и 1) имеет вид эллипса, что соответствует коррелированному нормальному распределению вида:

$$f(n_0, n_1) = \frac{1}{2\pi N \sigma_0 \sigma_1 \sqrt{1 - \rho^2}} \exp \left(\frac{-1}{2N(1 - \rho^2)} \left(\frac{(\Delta n_0)^2}{\sigma_0^2} + \frac{(\Delta n_1)^2}{\sigma_1^2} - 2\rho \frac{\Delta n_0 \Delta n_1}{\sigma_0 \sigma_1} \right) \right), \quad (5.5)$$

где $\Delta n_j = n_j - \mu_j N$, n_j — число вершин со степенью j , μ_j — средняя доля вершин, имеющих степень j , ρ — коэффициент корреляции, а σ_j^2 — дисперсии распределений, показанных на фиг. 3. Эти дисперсии оценены в формулах (5.4). Получим отсюда выражение для распределения $F(m)$ суммы случайных величин $m = n_0 + n_1$. Поскольку совместное распределение (5.5) степеней вершин известно, то имеем

$$F_m = \int_{-\infty}^{\infty} f(x, m - x) dx. \quad (5.6)$$

В силу быстрого убывания распределений (5.4) пределы интегрирования удобно распространить на всю числовую ось. В результате стандартного интегрирования плотностей гауссового типа получаем из (5.5), (5.6):

$$F(m) = \frac{1}{\sigma_0 \sigma_1 A \sqrt{2\pi N}} \exp \left(\frac{-(m - (\mu_0 + \mu_1)N)^2}{2A^2 N \sigma_0^2 \sigma_1^2} \right), \quad A^2 = \frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} + \frac{2\rho}{\sigma_0 \sigma_1}. \quad (5.7)$$

Если случайные величины n_0 и n_1 независимы, то дисперсия распределения (5.7) максимальна и равна $\sigma_{\max}^2 = \sigma_0^2 + \sigma_1^2$. Если же эти величины коррелированы, то дисперсия равна

$$\sigma_m^2 = A^2 \sigma_0^2 \sigma_1^2 = \sigma_0^2 + \sigma_1^2 + 2\rho\sigma_0\sigma_1. \quad (5.8)$$

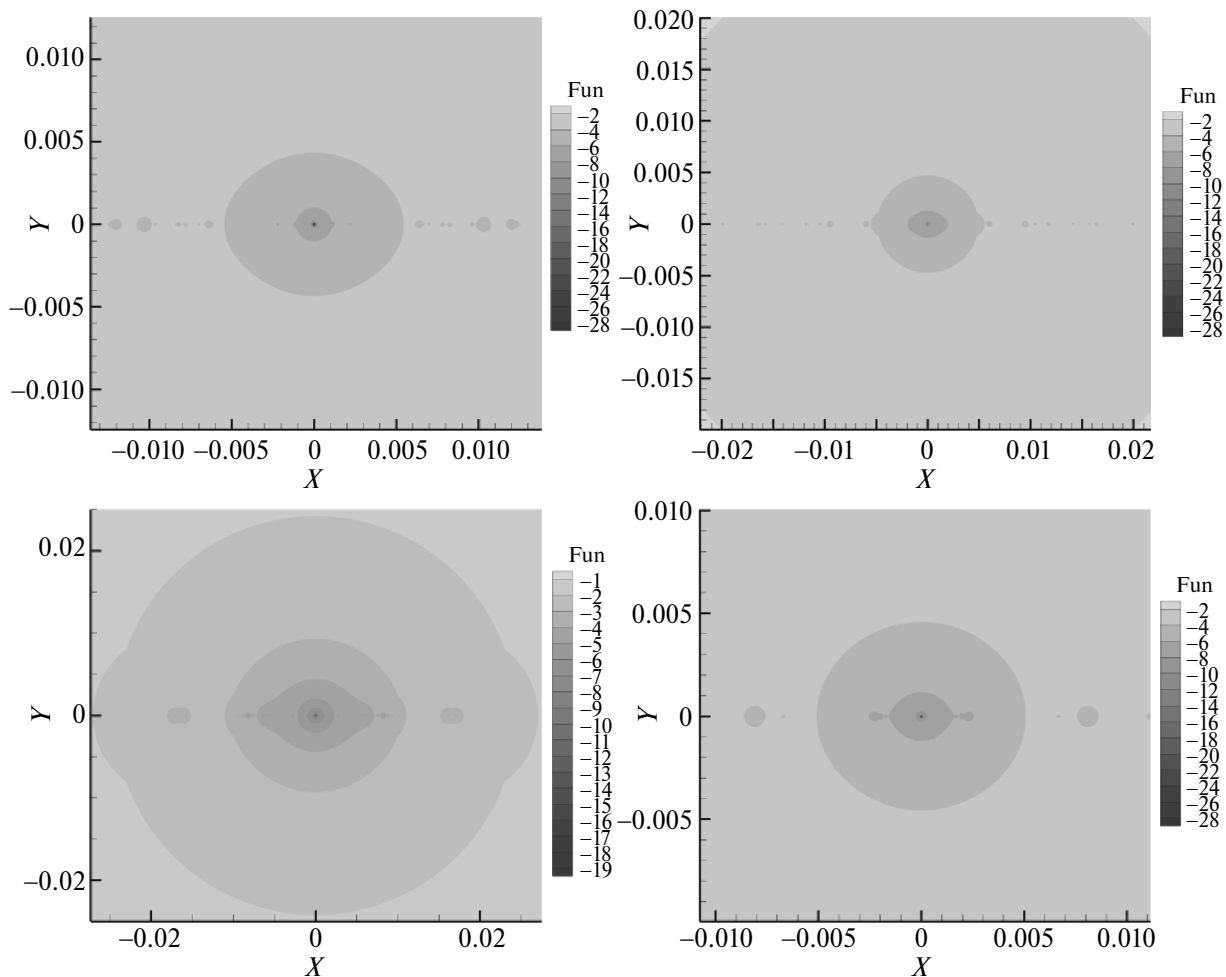
Минимальное значение дисперсии достигается при коэффициенте корреляции, равном $\rho = -1$, оно равно $\sigma_{\min}^2 = (\sigma_0 - \sigma_1)^2$. При равенстве дисперсий и линейной зависимости это распределение переходит в дельта-функцию.

По итогам вычислительного эксперимента были получены следующие числовые значения для параметров распределения (5.7):

$$\sigma_0^2 \approx 0.10; \quad \sigma_1^2 \approx 0.24; \quad \rho \approx -0.92. \quad (5.9)$$

Отсюда и из (5.8) находим, что дисперсия распределения (5.7) оказывается равной величине $\sigma_m^2 \approx 0.05$, что меньше, чем дисперсия распределения нулевых степеней. Следовательно, распределение суммы в данном случае имеет весьма малую дисперсию и может рассматриваться как приближенный закон сохранения средней суммы степеней.

Применим теперь к матрице инцидентности метод построения спектральных портретов, чтобы понять, где лежат точки спектра и с какой точностью требуется вычислять собственные значения. На фиг. 4 приведены фрагменты портретов матриц для случая 100×100 вблизи нулевых собственных значений. Точные ранги этих матриц равны соответственно 69, 64, 62 и 60. Границы областей, имеющих одну цветовую окраску, являются линиями уровня для L_2 -нормы резольвенты $\|(\lambda I - P)^{-1}\|$. Здесь P — матрица C_{ij}^{-1} из (5.2), элементы которой в данном случае определяются из (5.5). На легендах рисунков указаны показатели степени десятки, т.е. точности, с которой проведены расчеты спектра для данной цветовой окраски.



Фиг. 4. Фрагменты спектральных портретов четырех матриц инцидентности ГБС для $N = 100$.

Из вида ϵ -спектра этих матриц можно сделать следующие выводы. Во-первых, и это весьма важно, для более-менее корректного позиционирования нулевого собственного значения нужны вычисления с точностью на уровне $\epsilon = 10^{-10}$ и выше (в смысле величины отрицательной степени). С точностью же 10^{-4} спектр занимает некоторую область около нуля и отдельные точки не различимы. Во-вторых, с увеличением ранга (первые два портрета) спектр распознается более детально. Поэтому, с одной стороны, при численном анализе плохо обусловленных матриц, элементы которых известны точно, требуется достаточно высокая точность вычислений. В то же время эта точность, как показывают спектральные портреты, в практически важном случае не является недостижимой, а находится на вполне реализуемом уровне.

Таким образом, что касается ранга матриц рассматриваемого специального вида, то он может быть определен точно исходя из свойств ближайших соседей. Важно, что спектр матриц инцидентности, определяемый по методу Годунова, дает те же результаты: начиная с определенного уровня точности все ненулевые собственные значения идентифицируются как области спектрального портрета. Это означает, что вычислительные процедуры, проводимые с такими вырожденными матрицами, могут осуществляться корректно при надлежащей точности вычислений.

6. ЗАКЛЮЧЕНИЕ

В представленном исследовании приведено применение метода спектральных портретов плохо обусловленных матриц для статистического анализа структур графов на примере матриц инцидентности графов ближайших соседей. Показано, что точно решаемый пример с определением ранга матрицы инцидентности может быть верно решен и численно с использованием метода Годунова анализа ϵ -спектра матрицы. При этом корректно определяется не только ранг матрицы, что может быть определено и более простыми вычислительными процедурами, но и достаточно точное расположение собственных значений такой матрицы, что представляет самостоятельную ценность.

Результаты работы могут быть использованы для построения критерия независимости выборочных данных, поскольку найдено распределение суммы степеней вершин порядка 0 и 1 найдено с высокой статистической точностью, причем это распределение не зависит от распределения элементов в матрице расстояний, на основе которой проводится дальнейшее исследование.

Также следует указать о возможности применения данного метода, как и подхода с использованием спектральных портретов, к задачам в области случайных конечных автоматов.

СПИСОК ЛИТЕРАТУРЫ

1. Уилкинсон Дж.Х. Алгебраическая проблема собственных значений. М.: Наука, 1970. 564 с.
2. Голуб Дж., Ван Лоун Ч. Матричные вычисления. М.: Мир, 1999. 546 с.
3. Орлов Ю.Н., Осминин К.П. Методы статистического анализа литературных текстов. М.: Эдиториал УРСС/Книжный дом "ЛИБРОКОМ", 2012. 312 с.
4. Годунов С.К. Современные аспекты линейной алгебры. Новосибирск: Научн. книга, 1997. 388 с.
5. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Springer, 2001. 764 p.
6. Erdos P., Renyi A. On random graphs I // Publ. Math. Debrecen, 1959. Vol. 6. P. 290–297.
7. Колчин В.Ф. Случайные графы. М.: Физматлит, 2004. 256 с.
8. Bollobas B. Random Graphs. London: Cambridge University Press, 2001. 520 p.
9. Janson S., Luczak T., Rucinski A. Random graphs. New York: Wiley, 2000. 333 p.
10. Райгородский А.М. Модели случайных графов и их применения // Тр. МФТИ, 2010. Т. 2. № 4. С. 130–140.
11. Райгородский А.М. Модели Интернета. Долгопрудный: Издательский Дом Интеллект, 2013. 64 с.
12. Barabasi L.-A., Albert R. Emergence of scaling in random networks // Science. 1999. V. 286. P. 509–512.
13. Leskovec J., Chakrabarti D., Kleinberg J., Faloutsos C., Gharamani Z. Kronecker graphs: an approach to modeling networks // J. Machine Learning Research. 2010. V. 11. P. 985–1042.
14. Кислицын А.А. Исследование статистик графов ближайших соседей // Матем. моделирование. 2022. Т. 34. № 8. С. 110–126.

15. *Kislitsyn A.A., Orlov Yu.N.* Discussion about Properties of First Neighbor Graphs // Lobachevskii Journal of Mathematics 2022. V. 43. № 12. P. 109–118.
16. *Кислицын А.А.* Анализ каталога землетрясений // Матем. моделирование 2023. Т. 35. № 7. С. 63–82.
17. *Мельников С.Ю.* Идентификация конечных автоматов на основе метода многогранников. М. — Ижевск: Институт компьютерных технологий, 2013. 136 с.

ON SPECTRAL PORTRAITS OF NEAREST NEIGHBOR GRAPH INCIDENCE MATRICES

A.A. Kislitsyn*

Keldysh Institute of Applied Mathematics of the Russian Academy of Sciences, Miusskaya sq., 4, Moscow, 125047 Russia

**e-mail: alexey.kislitsyn@gmail.com*

Received 02 April, 2024

Revised 02 April, 2024

Accepted 05 May, 2024

Abstract. This paper explores the possibility of applying S.K. Godunov's method of constructing spectral portraits of matrices to estimate the rank of special types of matrices that appear in various applications, such as nearest neighbor graph structure analysis, finite automata theory, and sparse matrix spectrum estimation. A computational algorithm for generating an ensemble of random distance matrices and the associated nearest neighbor graphs is described. Based on computational experiments, parameters of the vertex degree distribution of random nearest neighbor graphs are evaluated. These estimates are feasible because the distribution is independent of the random distance function and follows a multivariate normal distribution. It is proven that the rank of the incidence matrix of a nearest neighbor graph equals the total number of vertices with in-degree 0 and 1, and the rank distribution of such a matrix is derived. It is also shown that, in this context, a method based on analyzing the vertex degree distribution is highly effective for determining the matrix rank.

Keywords: nearest neighbor graph, vertex degree distribution, spectral portrait, matrix rank distribution.