

МИНОБРНАУКИ РОССИИ

**Федеральное государственное бюджетное образовательное учреждение
высшего образования**

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра теоретических основ
компьютерной безопасности и
криптографии

Деревья и их приведенные колоды

АВТОРЕФЕРАТ

дипломной работы

студентки 6 курса 631 группы

специальности 10.05.01 Компьютерная безопасность

факультета компьютерных наук и информационных технологий

Володиной Полины Алексеевны

Научный руководитель

зав. кафедрой, д. ф.-м. н., доцент

23.01.2021 г.

М. Б. Абросимов

Заведующий кафедрой

д. ф.-м. н., доцент

23.01.2021 г.

М. Б. Абросимов

Саратов 2021

ВВЕДЕНИЕ

В современном мире большую роль играет информация, которая рассматривается как один из основных ресурсов информационного общества. Хранение, обработка и передача информации осуществляется с помощью информационных систем. Актуальной является проблема защиты информации в информационных системах, одним из главных этапов которой является описание защищаемой системы и алгоритма ее работы. Существует много различных способов описания и анализа информационных систем. К функциональному (структурному) методу также относится использование теории графов.

Высокие темпы информатизации различных видов деятельности в настоящее время привели к тому, что возникла необходимость компьютерного моделирования и проектирования сложных систем, изучения их свойств и управления ими в условиях дефицита времени, ограниченности ресурсов, неполноты информации. Но для исследования характеристик любой системы математическими методами должна быть обязательно выполнена формализация, то есть построена математическая модель системы. Графы являются хорошей математической моделью различных классов объектов и процессов.

Важным классом графов являются деревья. Многие задачи, весьма сложные в общем случае, эффективно решаются для деревьев. Одной из основных проблем в теории графов является проблема реконструкции графов по их частям.

В данной работе исследуется гипотеза реконструируемости Келли-Улама, а также некоторые доказанные частные случаи этой гипотезы. Далее рассматривается усиление теоремы Харари и Палмера об определимости дерева своей колодой, а именно вопрос об определимости дерева его приведенной колодой. Также в данной работе будет проведен анализ построенных приведенных колод для деревьев с количеством вершин $4 \leq n \leq 25$,

подсчитано количество деревьев с различными значениями размеров приведенных колод, построено распределение количества деревьев по заданным характеристикам, полностью описаны и классифицированы деревья с размером приведенной колоды 1, сделаны некоторые теоретические выводы на основе результатов вычислительных экспериментов, а также построено распределение количества многоуровневых звезд различных уровней.

КРАТКОЕ СОДЕРЖАНИЕ

В разделе 1 «Основные определения» приводятся основные теоретические понятия и определения^{1,2,3,4,5,6,7,8,9}, которые используются в данной работе.

Дерево – это связный граф, в котором нет циклов¹.

Два графа $G_1 = (V_1, \alpha_1)$ и $G_2 = (V_2, \alpha_2)$ называются изоморфными, если можно установить взаимно однозначное соответствие $f: V_1 \rightarrow V_2$, сохраняющее отношение смежности: $(u, v) \in \alpha_1$ тогда и только тогда, когда $(f(u), f(v)) \in \alpha_2$ для любых $u, v \in V_1$.

Подграф – граф, получающийся удалением произвольного количества вершин и всех смежных с ними ребер из исходного графа.

Максимальный подграф – подграф, получающийся удалением одной произвольной вершины.

Колодой графа называется список его максимальных подграфов⁴.

Поддерево – дерево, получающееся удалением произвольного количества висячих вершин и всех смежных с ними ребер из исходного дерева.

¹ Богомолов, А. М. Алгебраические основы теории дискретных систем / А. М. Богомолов, В. Н. Салий. – М. : Наука, Физматлит, 1997. – 368 с.

² Харари, Ф. Теория графов [Электронный ресурс] / Ф. Харари. – М. : Мир, 1973. – 300 с. – Загл. с экрана. – Яз. рус.

³ Абросимов, М. Б. О нижней оценке числа ребер 1-расширения сверхстройного дерева [Электронный ресурс] / М. Б. Абросимов // mathnet.ru [Электронный ресурс]. – URL: <http://www.mathnet.ru/links/f0d53351998eedb50dc567e6176926cb/isu258.pdf> (дата обращения: 01.12.2020). – Загл. с экрана. – Яз. рус.

⁴ Абросимов, М. Б. Практические задачи по графам: учебное пособие / М. Б. Абросимов, А. А. Долгов. – Саратов : Научная книга, 2009. – 75 с.

⁵ Касьянов, В. Н. Графы в программировании: обработка, визуализация и применение / В. Н. Касьянов, В. А. Евстигнеев. – СПб. : БХВ-Петербург, 2003. – 1104 с.

⁶ Knuth, D. E. The Art of Computer Programming / Donald E. Knuth. – 1997. – 650 p.

⁷ Оре, О. Теория графов [Электронный ресурс] / О. Оре. – М. : Наука, Физматлит, 1980. – 336 с. – Загл. с экрана. – Яз. рус.

⁸ Павлов, Д. А. Канонический код дерева [Электронный ресурс] / Д. А. Павлов // ipo.spb.ru [Электронный ресурс]. – URL: <http://ipo.spb.ru/journal/content/1037/Канонический%20код%20дерева.pdf> (дата обращения: 15.11.2020). – Загл. с экрана. – Яз. рус.

⁹ Read, R. C. The coding of various kinds of unlabeled trees [Электронный ресурс] / Ronald. C. Read // ScienceDirect [Электронный ресурс]. – 1972. – URL: <https://cutt.ly/XhXGWcH> (дата обращения: 14.11.2020). – Загл. с экрана. – Яз. англ.

Максимальное поддереву дерева – поддереву, получающееся удалением одной произвольной висячей вершины.

Колодой дерева будем называть список его максимальных поддеревьев.

Приведенная колода дерева – колода дерева, которая содержит только неизоморфные поддеревья дерева.

В разделе 2 «Деревья» рассказывается о значении деревьев в различных областях математики и других наук^{10,11}, а также приводится основная теорема о деревьях и алгоритм построения плоского изображения дерева¹.

В разделе 3 «Реконструкция деревьев» рассматривается гипотеза реконструируемости Келли-Улама^{12,13}, которая является одной из самых знаменитых открытых проблем в теории графов. Различным вариациям данной гипотезы посвящено много работ, например, работа Пола Келли¹⁴ и работа Фрэнка Харари и Эда Палмера¹⁵. Таким образом, известно, что дерево однозначно определяется его колодой¹⁵.

¹⁰ Особенности применения теории графов при решении задач и в практической деятельности [Электронный ресурс] // Studbooks.net [Электронный ресурс] : Студенческая библиотека онлайн. – URL: https://studbooks.net/2300871/matematika_himiya_fizika/grafy_derevyu_koren (дата обращения: 04.01.2021). – Загл. с экрана. – Яз. рус.

¹¹ Теория графов. Деревья. Сети. Сетевые модели представления информации. Применение графов и сетей [Электронный ресурс] // Учебно-методический кабинет [Электронный ресурс]. – URL: <https://ped-kopilka.ru/diskretnaja-matematika/teorija-grafov-derevja-seti-setevye-modeli-predstavlenija-informaci-primenenie-grafov-i-setei.html> (дата обращения: 04.01.2021). – Загл. с экрана. – Яз. рус.

¹² Kelly, P. J. On isometric transformations [Электронный ресурс] / P. J. Kelly // Minds@UW [Электронный ресурс]. – 1942. – URL: <https://minds.wisconsin.edu/bitstream/handle/1793/28652/38918336.pdf?sequence=1&isAllowed=y> (дата обращения: 05.01.2021). – Загл. с экрана. – Яз. англ.

¹³ Ulam, S. M. A collection of mathematical problems [Электронный ресурс] / S. M. Ulam // Hathi Trust [Электронный ресурс]. – 1960. – URL: <https://babel.hathitrust.org/cgi/pt?id=mdp.39015002923301&view=2up&seq=1> (дата обращения: 05.01.2021). – Загл. с экрана. – Яз. англ.

¹⁴ Kelly, Paul J. A congruence theorem for trees [Электронный ресурс] / Paul. J. Kelly // Project Euclid – mathematics and statistics online [Электронный ресурс]. – 1957. – URL: https://projecteuclid.org/download/pdf_1/euclid.pjm/1103043674 (дата обращения: 05.01.2021). – Загл. с экрана. – Яз. англ.

¹⁵ Harary, F. The reconstruction of a tree from its maximal subtrees [Электронный ресурс] / Frank Harary, Ed Palmer // Cambridge University Press – Canadian Journal of Mathematics [Электронный ресурс]. – 1966. – URL: https://www.cambridge.org/core/services/aop-cambridge-core/content/view/91BE62244C7322893AA5F05BE9DA61D6/S0008414X0004075Xa.pdf/reconstructi%20on_of_a_tree_from_its_maximal_subtrees.pdf (дата обращения: 05.01.2021). – Загл. с экрана. – Яз. англ.

В разделе 4 исследуется проблема изоморфизма деревьев, а также рассматривается вариант ее решения с помощью нахождения канонических кодов деревьев. Существует множество различных канонических кодов, в дипломной работе используется канонический код из книги «Алгоритмы на деревьях и графах»¹⁶, далее приводится описание данного канонического кода и пример его вычисления для некоторых деревьев.

В разделе 5 «Алгоритм поиска центра дерева» приводится теорема о центре дерева², а также алгоритм поиска центра дерева¹⁷, который используется в программной реализации.

В разделе 6 «Программная реализация» приводится описание и пример работы программной части дипломной работы.

На вход программе подается список деревьев в формате `sparse6`¹⁸, при этом все деревья из этого списка содержат одинаковое количество вершин. Списки деревьев генерируются с помощью программы *gentreeg*, входящей в пакет программ *nauty* and *Traces*¹⁹.

В результате работы программа делает вывод о том, однозначно ли определяются деревья с заданным количеством вершин своими приведенными колодами. Также в программе находится распределение количества деревьев по значениям размера приведенной колоды, частота встречаемости поддеревьев в приведенных колодах деревьев, распределение количества деревьев по парам значений: количество листьев (размер колоды дерева) и размер приведенной колоды. Данная статистика записывается в виде таблиц в файлы в формате `.txt`.

¹⁶ Valiente, G. Algorithms on Trees and Graphs [Электронный ресурс] / G. Valiente // World Documents [Электронный ресурс]. – 2002. – URL: <https://vdocuments.net/algorithms-on-trees-and-graphs-.html> (дата обращения: 01.10.2020). – Загл. с экрана. – Яз. англ.

¹⁷ Алгоритмы на деревьях [Электронный ресурс] // Университет ИТМО [Электронный ресурс]. – URL: https://neerc.ifmo.ru/wiki/index.php?title=Алгоритмы_на_деревьях (дата обращения: 05.10.2020). – Загл. с экрана. – Яз. рус.

¹⁸ Description of graph6, sparse6 and digraph6 encodings [Электронный ресурс] // users.cecs.anu.edu.au [Электронный ресурс]. – URL: <https://users.cecs.anu.edu.au/~bdm/data/formats.txt> (дата обращения: 17.09.2020). – Загл. с экрана. – Яз. англ.

¹⁹ McKay, B. D. Practical Graph Isomorphism, II [Электронный ресурс] / B. D. McKay, A. Piperno // ScienceDirect [Электронный ресурс]. – 2014. – URL: <https://cutt.ly/hyFu7Qz> (дата обращения: 16.09.2020). – Загл. с экрана. – Яз. англ.

Листинг программы приведен в Приложении А дипломной работы.

Раздел 7 «Анализ результатов вычислительных экспериментов» включает в себя различную статистическую информацию, полученную в результате работы программы, и выводы, которые были сделаны на основе анализа этой информации.

В Приложении Б представлено распределение количества деревьев по парам значений: размер колоды (количество листьев) и размер приведенной колоды для деревьев с количеством вершин $4 \leq n \leq 25$.

В подразделе 7.1 в таблице 4 приведены полученные выводы о том, однозначно ли определяются деревья своими приведенными колодами. На их основе была сформулирована следующая гипотеза.

Гипотеза. Все деревья с количеством вершин $n \geq 7$ однозначно определяются своей приведенной колодой.

В таблице 2 представлено распределение количества деревьев по значениям размера приведенной колоды для деревьев с количеством вершин $4 \leq n \leq 25$. На основе этой информации был получен теоретический результат о максимальном размере приведенной колоды для заданного количества вершин n , сформулированный в подразделе 7.2.

Далее в подразделе 7.3 приведено описание и подсчет количества деревьев с количеством листьев (размером колоды) $n - 2$ и размером приведенной колоды 2 (n – количество вершин).

1) Описание.

Деревья с количеством листьев (размером колоды) $n - 2$ и размером приведенной колоды 2 выглядят следующим образом: ствол, представленный в виде цепи P_2 , из концевых вершин ствола выходят цепи длины 1, то есть вершины, которые смежны с концевыми вершинами ствола, являются листьями дерева и их количество будет равно $n - 2$. Обязательным условием является то, что количество цепей, выходящих из одной концевой вершины ствола, не равно

количеству цепей, выходящих из другой концевой вершины ствола, потому что в данном случае размер приведенной колоды дерева был бы равен 1.

На рисунке 1 представлены деревья с количеством листьев (размером колоды) $n - 2$ и размером приведенной колоды 2, где $5 \leq n \leq 10$.

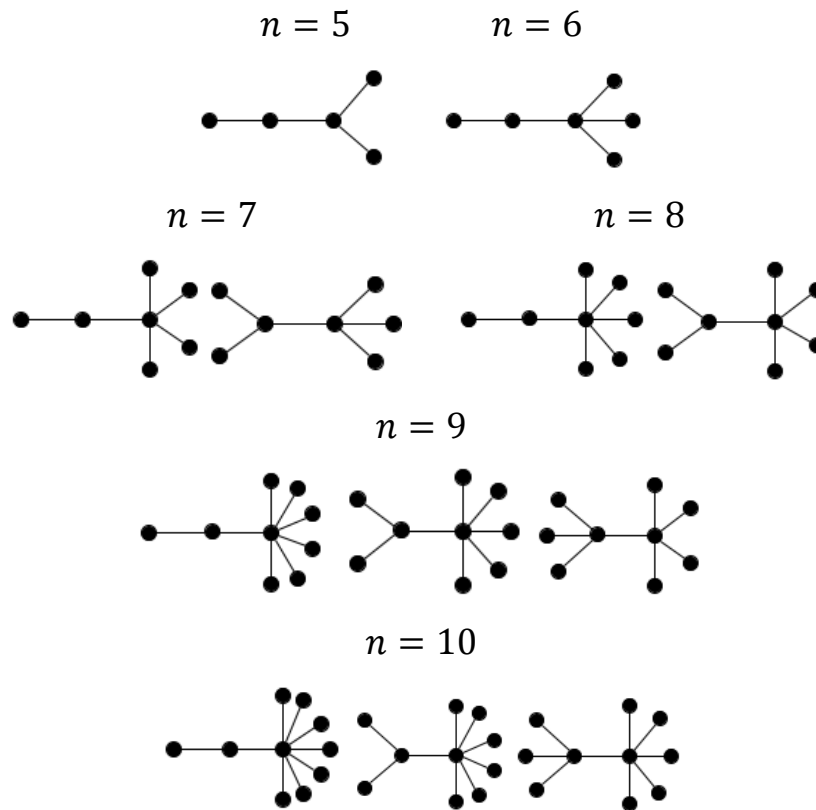


Рисунок 1 – Деревья с количеством листьев (размером колоды) $n - 2$ и размером приведенной колоды 2, где $5 \leq n \leq 10$

2) Подсчет количества.

В описанных выше деревьях количество листьев равно $n - 2$, при этом все листовые вершины смежны с одной из двух концевых вершин ствола такого дерева. Таким образом, чтобы найти количество описанных выше деревьев, зная только количество вершин n , нужно найти количество разложений числа $n - 2$ на два неравных между собой слагаемых. Обозначим количество таких деревьев через $N_{n-2,2}$. Получаем следующую формулу:

$$\begin{cases} N_{n-2,2} = \frac{n-2}{2} - 1, \text{ если } n - \text{ четное} \\ N_{n-2,2} = \left\lfloor \frac{n-2}{2} \right\rfloor, \text{ если } n - \text{ нечетное} \end{cases}$$

В случае четного n находим количество разложений числа $n - 2$ на два слагаемых и вычитаем 1, так как для четного n в это количество будет входить разложение числа $n - 2$ на два равных между собой слагаемых.

Подраздел 7.4 включает в себя описание деревьев, которые реже всего встречаются в приведенных колодах.

В разделе 8 «Классификация деревьев с размером приведенной колоды 1» были рассмотрены и полностью описаны деревья с размером приведенной колоды 1, а также предложена их классификация и приведен алгоритм подсчета количества таких деревьев без проведения каких-либо вычислительных экспериментов.

Введем обозначение для деревьев с размером приведенной колоды 1: SC (Star Central). Будем называть такие деревья многоуровневыми звездами. Каждый уровень кодируется двумя параметрами – числом разветвлений и длиной ствола, при этом число разветвлений должно быть > 1 , ствол считается от ближайшей центральной вершины. Бицентральные многоуровневые звезды будем обозначать SB (Star Bicentral).

Многоуровневые звезды первого уровня кодируются двумя параметрами и обозначаются SC_{n_1, n_2} или SB_{n_1, n_2} . Число вершин n при этом находится по следующей формуле:

- $n = 1 + n_1 n_2$, если дерево является центральным;
- $n = 2 + n_1 n_2$, если дерево является бицентральным.

Многоуровневые звезды второго уровня кодируются четырьмя параметрами и обозначаются SC_{n_1, n_2, n_3, n_4} или SB_{n_1, n_2, n_3, n_4} . Число вершин n :

- $n = 1 + n_1 (n_2 + n_3 n_4)$, если дерево является центральным;
- $n = 2 + n_1 (n_2 + n_3 n_4)$, если дерево является бицентральным.

Многоуровневые звезды последующих уровней задаются рекурсивно аналогичным образом. Также важно отметить, что в случае бицентральных деревьев параметр n_1 должен быть четным.

Пусть n – количество вершин дерева.

Первая многоуровневая звезда первого уровня появляется при $n = 3$: $SC_{2,1}$; второго уровня – при $n = 7$: $SC_{2,1,2,1}$; третьего уровня – при $n = 15$: $SC_{2,1,2,1,2,1}$; четвертого уровня – при $n = 31$: $SC_{2,1,2,1,2,1,2,1}$.

Таким образом, каждая первая многоуровневая звезда k -ого уровня появляется при $n = 2^{k+1} - 1$.

Для того, чтобы найти количество деревьев с размером приведенной колоды 1, зная только количество вершин n , необходимо перебирать подходящие параметры $n_1, n_2, \dots$. Формула для количества вершин n в общем случае будет выглядеть так: $n = 1$ (или 2) + $n_1(n_2 + \dots + n_{m-1}n_m)$, где $m = 2k$, k – наибольший возможный уровень для данного n .

Общее количество таких деревьев будет равно сумме количеств звезд разных уровней, в случае нечетного количества вершин перебираем подходящие центральные деревья, а в случае четного количества вершин еще и бицентральные.

В таблице 5 представлено распределение количества многоуровневых звезд разных уровней для деревьев с количеством вершин $5 \leq n \leq 25$. На основе информации, представленной в этой таблице, были сделаны некоторые выводы:

1. Для нечетного n (n – количество вершин дерева): количество многоуровневых звезд 1 уровня равно количеству разложений числа $n - 1$ на два множителя, при этом первый множитель должен быть > 1 .

2. Для четного n : количество центральных многоуровневых звезд 1 уровня равно количеству разложений числа $n - 1$ на два множителя, при этом первый множитель должен быть > 1 ; количество бицентральных многоуровневых звезд 1 уровня равно количеству разложений числа $n - 2$ на два множителя, при этом первый множитель должен быть четным и > 1 .

3. Для четного n : если количество центральных многоуровневых звезд 1 уровня $= 1$, то для данного n не будет существовать центральных многоуровневых звезд уровней, больших чем 1.

На рисунке 2 для примера представлены деревья с количеством вершин $n = 16$ и размером приведенной колоды 1 с соответствующими им введенными обозначениями.

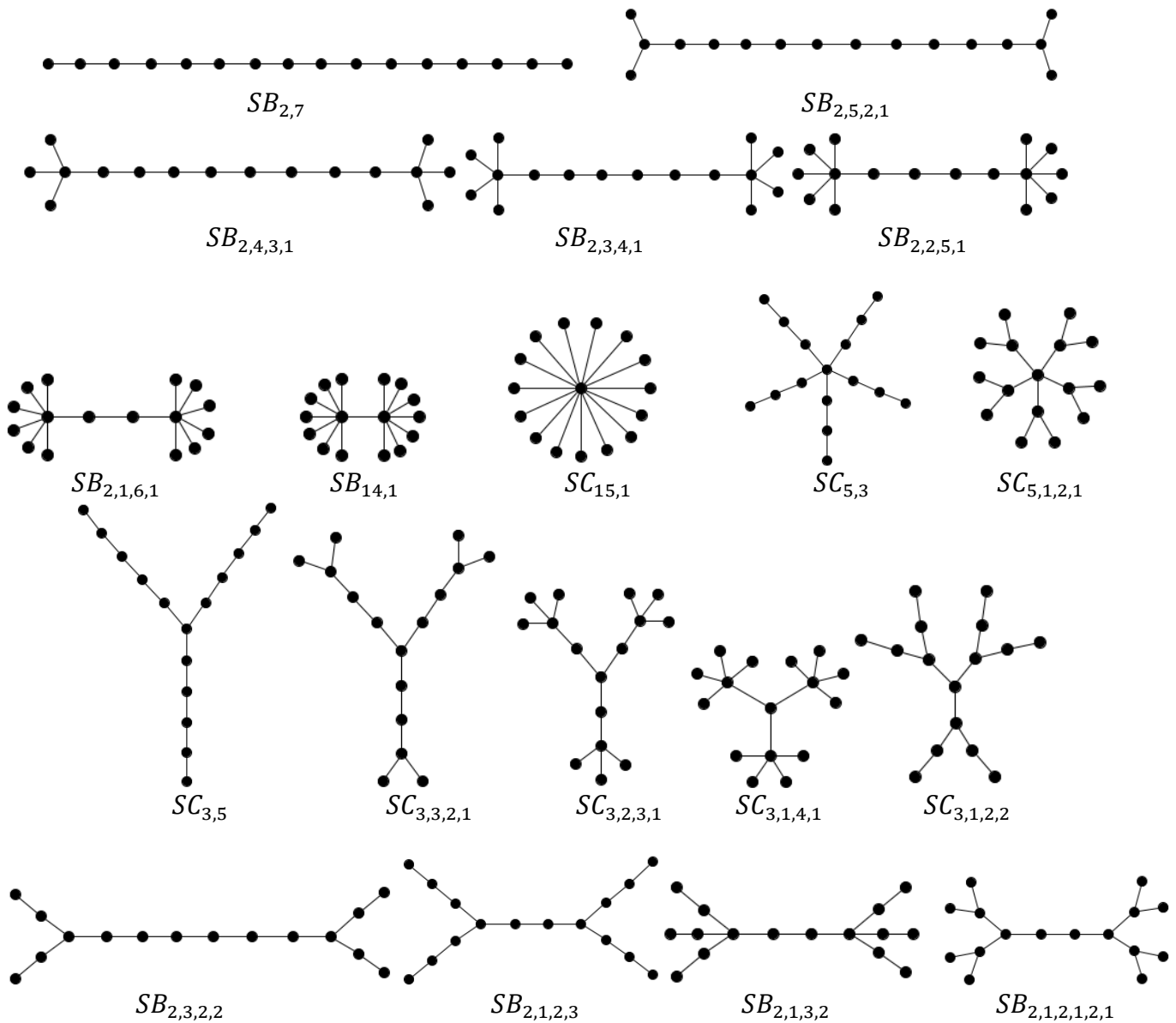


Рисунок 2 – Деревья с количеством вершин $n = 16$ и размером приведенной колоды 1

ЗАКЛЮЧЕНИЕ

В современном мире теория графов используется при проектировании интегральных схем, схем управления, ЭВМ и баз данных, при разработке вычислительных и информационных сетей. Деревья широко распространены в теории электрических цепей, химии, вычислительной технике и в информатике.

В ходе написания данной работы была рассмотрена теория о способах решения проблемы изоморфизма деревьев, на основе которой была написана программа на языке Python, реализующая построение приведенной колоды дерева. Программа также строит распределение количества деревьев по различным заданным характеристикам, выдает результат о том, определяются ли деревья с количеством вершин $4 \leq n \leq 25$ своими приведенными колодами. На основе анализа результатов, полученных в ходе вычислительных экспериментов, были сформулированы некоторые теоретические выводы, а также предложена классификация деревьев с размером приведенной колоды 1 и приведено распределение количества многоуровневых звезд различных уровней.