

Министерство образования и науки Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «САРАТОВСКИЙ  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ  
УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра теоретических основ  
компьютерной безопасности и  
криптографии

### **Data mining в теории графов**

АВТОРЕФЕРАТ

дипломной работы

студентки 6 курса 631 группы

специальности 10.05.01 Компьютерная безопасность

факультета компьютерных наук и информационных технологий

Фоминой Александры Викторовны

Научный руководитель

профессор, д.ф.-м.н.

М.Б. Абросимов

Заведующий кафедрой

профессор, к.ф.-м.н.

В.Н. Салий

Саратов 2017

## ВВЕДЕНИЕ

*Инвариантом* графа называется набор его характеристик, одинаковых для всех изоморфных ему графов. Инвариантами графа являются, например, число вершин графа, количество ребер.

На данный момент известны некоторые зависимости между инвариантами графов, например, что число вершинной связности не превосходит числа реберной связности, или что хроматическое число не меньше частного от деления числа вершин графа и числа независимости.

В то же время проблема поиска и выявления других интересных связей между инвариантами графа остается актуальной и сейчас.

Данная работа была вдохновлена программой Grinvin, которая служит для изучения графов, их инвариантов и отношений между ними. Она позволяет строить гипотезы о выбранном множестве инвариантов и генерирует контрпример – граф, для которого данная гипотеза не верна.

Также на создание дипломной работы повлиял веб-портал The House of Graphs. Он содержит базу данных интересных графов с возможностью поиска по инвариантам.

Целью работы является изучение возможности нахождения ранее не известных зависимостей между различными инвариантами графа.

В ходе работы была сгенерирована база данных всех графов с числом вершин до 10. Общее число графов в исследовании: 15153159.

Реализованы алгоритмы вычисления таких инвариантов графа, как связность и количество компонент связности, обхват, двудольность, реберная связность, вершинная связность, радиус, диаметр, примитивность и экспонент, хроматическое число, число независимости. Эти инварианты были вычислены для графов и сохранены в базу данных.

Была создана программа с веб-интерфейсом, позволяющая производить поиск графов по выбранным инвариантам, а также позволяющая строить и проверять гипотезы насчет инвариантов на заданном множестве графов.

Найденные гипотезы для связных графов представлены в приложении Б, для всех графов – в приложении В.

Дипломная работа состоит из введения, 4 разделов, заключения, списка использованных источников и 3 приложений. Общий объем работы – 76 страниц, из них 35 страниц – основное содержание, включая 11 рисунков и 1 таблицу, список использованных источников из 13 наименований.

## КРАТКОЕ СОДЕРЖАНИЕ

Раздел 1 «Основные понятия» содержит основные термины и определения из теории графов, которые используются для вычисления инвариантов в работе.

В разделе 2 «Инварианты графа и их нахождение» определены понятия инвариантов, рассматриваемых в работе, приведены алгоритмы их нахождения и сложности вычисления алгоритмов. В пункте 2.1 предложен способ проверки графа на связность и нахождение количества компонент связности с помощью алгоритма поиска в глубину. Пункт 2.2 рассказывает о том, как вычислять обхват графа также поиском в глубину. Далее, в пункте 2.3 представлен алгоритм проверки графа на двудольность поиском в ширину.

Задача нахождения реберной связности сводится к задаче поиска минимального глобального разреза, который можно вычислить алгоритмом Штор-Вагнера. Его описание находится в пункте 2.4.1.

Для определения вершинной связности используется теорема Менгера для вершинной  $k$ -связности (п. 2.5.1). Она говорит о том, что наименьшее число вершин, разделяющих две несмежные вершины  $u$  и  $v$ , равно наибольшему числу непересекающихся простых  $(u - v)$  цепей. Отсюда следует, что граф  $G$  является вершинно  $k$ -связным тогда и только тогда, когда любая пара его вершин соединена по крайней мере  $k$  вершинно непересекающимися путями. Таким образом, задача нахождения вершинной связности сводится к нахождению вершинно-непересекающихся путей. Если преобразовать граф, разбив каждую вершину  $v$  графа на две вершины  $v_1$  и  $v_2$ , все ребра, которые входили в  $v$  преобразовать во входящие в  $v_1$ , и все ребра, которые выходили из  $v$  преобразовать в выходящие из  $v_2$ , а также добавить ребро  $(v_1, v_2)$  с пропускной способностью 1, можно свести задачу к нахождению реберной связности в новом графе.

Для нахождения радиуса и диаметра графа можно найти расстояния между каждой парой вершин, воспользовавшись алгоритмом Флойда-Варшалла (п. 2.6.1). После этого для каждой вершины найти максимальное из расстояний от

нее до всех остальных вершин, тем самым получив эксцентриситеты вершин. В итоге получим радиус и диаметр как минимальный и максимальный из полученных эксцентриситетов соответственно.

Для проверки графа на примитивность и нахождения его экспонента (п. 2.7) используется следствие из определения примитивного графа, а именно:

Если граф примитивный с экспонентом  $e$ , то это равносильно тому, что в матрице смежности графа  $A^e$  все элементы равны единице.

На основе этого следствия, а также используя утверждение, что экспонент  $\text{exp}(G)$  графа  $G$  удовлетворяет неравенству  $\text{exp}(G) \leq n^2 - 2n + 2$ , предложен алгоритм нахождения экспонента, суть которого заключается в умножении матрицы смежности графа на себя, пока все элементы матрицы не будут равны единице, или пока экспонент удовлетворяет неравенству.

В пункте 2.8. приведено описание алгоритма нахождения хроматического числа графа, который работает на основе битовых операций над матрицей смежности.

Задача нахождения числа независимости, или задача о независимом множестве, принадлежит к классу  $NP$ -полных задач. Стратегия перебора для вычисления числа независимости и его вычислительная сложность представлены в п. 2.9.

Раздел дипломной работы 3 «Практическая часть» содержит сведения о генерации базы данных графов и разработанной программе. В пункте 3.1 описан способ генерации всевозможных графов с заданным числом вершин с помощью утилиты `geng` из пакета `nauty`. Изложен метод кодирования графов для хранения в базе данных.

В пункте 3.2 содержится описание разработанной программы, скриншоты и результаты работы. Список инвариантов, вычисляемых этой программой, находится в пункте 3.2.1. В пункте 3.2.2 приведены примеры запросов поиска в системе графов с различными параметрами.

В пункте 3.2.3 «Data mining» представлено описание процесса генерации гипотез. Приведены примеры сгенерированных гипотез на заданных

инвариантах. Полный список гипотез, полученных с помощью программы, для связных графов – находится в приложении Б, для всех графов – в приложении В.

В разделе 4 рассматриваются существующие аналоги разработанной программы. Пункт 4.1 рассказывает об отличиях веб-портала The House of Graphs, также имеющего базу данных графов с вычисленными инвариантами, от программы с веб-интерфейсом, созданной в рамках данной работы. В пункте 4.2 содержится описание программы Grinvin, которая также позволяет генерировать гипотезы для заданного множества графов.

## ЗАКЛЮЧЕНИЕ

В ходе выполнения дипломной работы были реализованы поставленные цели:

- создана база данных графов с числом вершин до 10 с помощью программного комплекса nauty;
- для графов подсчитаны выбранные инварианты: связность и количество компонент связности, обхват, двудольность, реберная связность, вершинная связность, радиус, диаметр, примитивность и экспонент, хроматическое число, число независимости;
- разработана программа с веб-интерфейсом, с помощью которого можно делать запросы поиска графов или поиска зависимостей между какими-либо инвариантами графов;
- с помощью разработанной программы определены зависимости между выбранными инвариантами графов, для связных графов и для всех графов, которые представлены в приложениях Б и В соответственно;
- были рассмотрены аналоги системы – веб-портал The House of Graph и программа Grinvin.

Разработанная программа написана на языке программирования Java 8 в среде IntelliJ IDEA 14.1.2.

Главная поставленная цель – найти ранее неизвестные зависимости между инвариантами графов – была выполнена. Найденные зависимости могут использоваться в дальнейших исследованиях связей выбранных инвариантов графов. Разработанная программа может применяться для изучения графов и их свойств, для поиска и выборки графов с нужными свойствами. Предложенные алгоритмы вычисления инвариантов не всегда были эффективными. Вычислительная сложность нахождения экспонента по приведенному в работе алгоритму составляет  $O(n^5)$ , но стоит уточнить, что в процессе выполнения дипломной работы не было обнаружено более эффективного алгоритма в каких-либо источниках. Вычисление числа независимости имеет экспоненциальную

сложность, т.к. эта задача принадлежит к классу  $NP$ -полных задач и на момент написания этой работы не доказано существования эффективного алгоритма ее решения.



## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Абросимов, М. Б. Практические задания по графам: учеб. пособие / М. Б. Абросимов, А. А. Долгов. 2-е изд. Саратов: Изд-во «Научная книга», 2009.
- 2 Алгоритмы: построение и анализ / Т. Кормен [и др]. 2-е изд. М.: «Вильямс», 2006.
- 3 Иванов Максим [Электронный ресурс] // MAXimal :: algo [Электронный ресурс]. URL: <http://e-maxx.ru/algo/> (дата обращения: 23.12.2016). Загл. с экрана. Яз. рус.
- 4 Комоско, Л. Ф. Быстрый алгоритм для решения задачи о раскраске графа с использованием битовых операций [Электронный ресурс] / Л. Ф. Комоско, М. В. Бацын. URL: <http://itas2014.iitp.ru/pdf/1569941697.pdf> (дата обращения: 23.12.2016). Яз. рус.
- 5 НОУ ИНТУИТ Лекция «Категории информационных систем» [Электронный ресурс] // НОУ «ИНТУИТ» [Электронный ресурс]. URL: <http://www.intuit.ru/studies/courses/1055/271/lecture/6876?page=7> (дата обращения: 24.12.2016). Загл. с экрана. Яз. рус.
- 6 Оре, О. Теория графов / О. Оре. М.: Наука, 1980.
- 7 Салий, В. Н. Минимальные примитивные расширения ориентированных графов / В. Н. Салий // Прикладная дискретная математика. 2008. №1(1), с. 116–119.
- 8 Фомичев, В. М. Новая универсальная оценка экспонентов графов / В. М. Фомичев // Прикладная дискретная математика. 2014. №2(24).
- 9 Харари, Ф. Теория графов / Ф. Харари. М.: УРСС, 2003.
- 10 Grinvin [Электронный ресурс]. URL: <http://www.grinvin.org/en/> (дата обращения: 23.12.2016). Загл. с экрана. Яз. англ.
- 11 House of Graphs [Электронный ресурс] / G. Brinkmann, K. Coolsaet, J. Goedgebeur, H. Mélot // House of Graphs [Электронный ресурс] : a database of interesting graphs, Discrete Applied Mathematics. 2013.

URL: <https://hog.grinvin.org/> (дата обращения: 23.12.2016). Загл. с экрана.

Яз. англ.

12 Stoer, M. A simple min-cut algorithm / M. Stoer, F. Wagner // Journal of the ACM. 1997. Vol. 44, No. 4. P. 585–591.

13 The on-line encyclopedia of integer sequences (OEIS) [Электронный ресурс].

URL: <https://oeis.org/> (дата обращения: 24.12.2016). Загл. с экрана. Яз. англ.