

Федеральное государственное автономное
образовательное учреждение высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт математики и фундаментальной информатики
Кафедра высшей и прикладной математики

УТВЕРЖДАЮ

Заведующий кафедрой

_____ / С.Г. Мысливец

«____» _____ 2021 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

ИССЛЕДОВАНИЕ ЗАДАЧ КЛАСТЕРИЗАЦИИ ГРАФА

Направление 01.04.02 Прикладная математика и информатика

Магистерская программа 01.04.02.06 Прикладная математика и информатика
в гуманитарных и социально-экономических науках

Научный руководитель

кандидат физико-математических наук,

доцент

_____/Д.В. Семенова

Выпускник

_____/М.К. Кумарбекова

Красноярск 2021

АННОТАЦИЯ

Целью работы является описание и изучение традиционных подходов к решению задач кластеризации графа.

Магистерская диссертация содержит введение, три главы, заключение, список использованных источников. Во введении описана актуальность исследования по выбранной теме, ставится цель и задачи исследования. В первой главе изложены элементы теории графов, даны определения, используемые в исследовании, а также введена общая постановка задачи кластеризации графа. Во второй главе рассмотрены алгоритмы решения задачи кластеризации графа. Приведенные алгоритмы кластеризации графа детально сопоставлены и подробно рассмотрены с примерами. Третья глава посвящена вычислительным экспериментам. В заключении даны основные выводы.

Ключевые слова: теория графов, кластер, задача кластеризации графа, алгоритмы кластеризации графа.

ABSTRACT

The aim of this work is to study approaches to solving the problem of cluster analysis of a graph.

The master's thesis includes an introduction, three chapters, a conclusion, a list of sources used. The introduction reveals the relevance of research in the chosen direction, sets the goal and objectives of the research. In the first chapter, the elements of graph theory are presented, definitions are given, used in research, and a general formulation of the graph clustering problem is introduced. In the second problem, algorithms for solving graph clustering are considered. The given graph clustering algorithms are compared in detail and considered in detail with examples. The third chapter is devoted to computational experiments. The conclusion provides the main findings.

Keywords: graph theory, cluster, graph clustering problems, graph clustering algorithms.

СОДЕРЖАНИЕ

Введение	3
1 Постановка задачи кластеризации графа	5
1.1 Основные определения и обозначения	6
1.2 Общая постановка задачи кластеризации графа	8
1.3 Выводы по первой главе	8
2 Описание алгоритмов	9
2.1 Задачи кластеризации графа, использующие метрики	9
2.1.1 Пример решения задач GC , GC_2 , $GC_{\leq 3}$	10
2.2 Задачи иерархической кластеризации графа	15
2.3 Задача спектральной кластеризации графа	18
2.4 Задача кластеризации графа, основанные на модулярности	21
2.5 Выводы по второй главе	22
3 Вычислительные эксперименты	24
3.1 Способы машинного представления графов	24
3.2 Вычислительные эксперименты	25
3.3 Выводы по третьей главе	26
Заключение	27
Список использованных источников	28

ВВЕДЕНИЕ

Актуальность работы. Изучение графов является актуальной задачей, поскольку на практике графы часто возникают при упрощении сложных систем. К примеру в виде графа отображают:

- взаимосвязи различных сайтов в интернете,
- социальные сети (сети контактов),
- генные сети в молекулярной биологии,
- порты, аэропорты, города (в качестве узлов графа) и соединяющие пути (в качестве ребер).

При работе с графами часто бывает полезно обнаружить в нем какую-нибудь структуру, например, группы вершин, внутри которых связей много, а между группами – мало. На практике сталкиваемся с большими графами, и для определения групп вершин, необходима помощь компьютера. Поэтому и создаются огромное количество алгоритмов, обладающих всевозможными свойствами.

Обзор литературы. В связи с растущим интересом исследователей к задачам кластеризации графа за последние годы было написано немало статей, освещающих разные аспекты данной темы. Граф – конечный набор объектов произвольной природы, называемых вершинами, и связей между некоторыми вершинами, называемых ребрами. В работе [8] собраны основные термины и основополагающие понятия и утверждения теории графов. Кластеризация графа [1, 15] – это задача о нахождении таких кластеров вершин, имеющих мало ребер между группами, и много ребер внутри кластеров. В литературе рассмотрены несколько задач кластеризации графа [35, 19]. Например, в обзорной статье, о задач кластеризации графа [9, 18, 33], рассматриваются три задачи кластеризации графа и предложены приближенные алгоритмы. В статье [26, 27] обсуждается более 15 алгоритмов, которые классифицированы по нескольким группам. Наборы данных из разных областей и имеющие разный смысл для сравнения алгоритмов, можно найти на следующих сайтах [30, 31, 32].

Цели и задачи. Целью работы является исследование задач кластеризации графа и анализ алгоритмов их решения. Достижение поставленной цели реализуется путем решения следующих задач.

1. Провести обзор литературы и изучить задачи кластеризации графа.
2. Исследовать существующие постановки задач кластеризации графа и алгоритмы их решения.
3. Провести вычислительные эксперименты.

Структура диссертации. Объём диссертационной работы 30 страниц, на которых размещены 13 рисунков и 1 таблица. При написании работы использовалось 39 источников.

ЗАКЛЮЧЕНИЕ

Основная цель диссертации – представить базовую идею традиционных, обычно используемых алгоритмов кластеризации, проанализировать преимущества и недостатки каждого из них. Представить полный список всех существующих алгоритмов кластеризации из-за разнообразия подходов, информации, пересечения областей исследований достаточно сложно. Прделанный анализ позволяет легко выбрать подходящую группу алгоритмов исходя из технической задачи.

В диссертационной работе получены следующие основные результаты.

- Сделан обзор существующих подходов к решению задач кластеризации графа.
- Исследованы шесть постановок задач кластеризации.
- Исследованы шесть алгоритмов кластеризации графа и приведены примеры их работы.
- Для спектрального алгоритма и алгоритма Гирвана–Ньюмена проведены вычислительные эксперименты, с использованием библиотек Python.

Работа докладывалась на научном семинаре кафедры Высшей и прикладной математики 2021.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Айвазян, С. А. Прикладная статистика : Классификация и снижение размерности / С. А. Айвазян, В. М. Бухштабер, И. С. Енюков, Л. Д. Мешалкин – Москва : Финансы и статистика, 1989. – 450 с.
2. Быкова, В. В. Комбинаторные алгоритмы: множества, графы, коды : учебное пособие для студентов вузов, обучающихся по направлению "Математика и компьютерные науки"/ В. В. Быкова ; Сиб. федер. ун-т, Ин-т математики и фундамент. информатики. - Красноярск : СФУ, 2015. - 150 с. : ил. - Библиогр.: с. 120-121. - 500 экз. - ISBN 978-5-7638-3155-9 : 160.00 р. - Изд. № 2015-14.
3. Волков, К. Н. Балансировка нагрузки процессоров при решении краевых задач механики жидкости и газа сеточными методами, Вычислительные методы и программирование. 2012.
4. Воронцов, К.В. Алгоритмы кластеризации и многомерного шкалирования. Курс лекций. МГУ, 2007.
5. Давыдов, О. А., Анализ существующих алгоритмов кластеризации (часть 1). УДК 681.518.5, Вестник ТОГУ, 2020.
6. Деза, Е. Энциклопедический словарь расстояний / Е. Деза, М. М. Деза. – Москва : Наука, 2008. – 448 с.
7. Дюран, Б. Кластерный анализ / Б. Дюран, П. Оделл. – М.: «Статистика», – 128 с.
8. Зыков А.А. Основы теории графов. - М: Вузовская книга, 2004. - 664 с.
9. Ильев, В. П., Ильева, С. Д., Навроцкая, А. А. Приближенные алгоритмы для задач аппроксимации графов, дискретный анализ и исследование операций // Январь—февраль 2011. Том 18, № 1. С. 41–60 УДК 519.8.
10. Касьянов, В.Н. Графы в программировании: обработка, визуализация и применение / Касьянов В.Н., Евстигнеев В.А. // СПб, 2003. — 1104 с.
11. Классификация и кластер. Под ред. Дж. Вэн Райзина. М.: Мир, 1980. 390 с.
12. Ламберто Гарсия дель Сид. Замечательные числа : Ноль, 666 и другие бес- тии. — М. : «Де Агостини», 2014. — Т. 21. — 160 с. — (Мир математики: в 40 т.). — ББК 22.1. — УДК 51(0.062)(G). — ISBN 978-5-9774-0682-6.

13. Мартынов, Н. Н. Кластеризация графов на основе оценок изменения модулярности // Н. Н. Мартынов, О. В. Хандарова, Ф. В. Хандаров, Известия ИГУ. Математика, 2018, т. 25, 63–78.
14. Спектральная кластеризация // Основа и применение William Fleshman, 2019 [Электронный ресурс]. <https://towardsdatascience.com/spectral-clustering-aba2640c0d5b>.
15. Силин, И., Обзор и экспериментальное сравнение алгоритмов кластеризации графов. / Силин, И., Панов, М., МФТИ.
16. Статья о промежуточности рёбер на AlgoWiki <http://algowiki\protect\discretionary{\char\hyphenchar\font}{}{}project.org/ru/>
17. Орлов, А. О., Чеповский, А. А. О свойствах модулярности и актуальных корректировках алгоритма Блонделя // Национальный исследовательский университет – Высшая школа экономики ул. Мясницкая, 20, Москва, Россия.
18. Фридман Г.Ш. Одна задача аппроксимации графов // Управляемые системы. – 1971. – Вып. 8. – С. 73–75.
19. Brandes U., Delling D., Gaertler M., Goerke R., Hoefer M., Nikoloski Z., Wagner D. Maximizing Modularity is hard // arXiv:physics/0608255. 2006.
20. Bansal N., Blum A., Chawla S. Correlation clustering // Machine Learning. 2004. V. 56. P. 89–113.
21. Chen, M. Community Detection via Maximization of Modularity and Its Variants. // K.Kuzmin, M.Chen. 2014
22. Clustering Based on Eigenvectors of the Adjacency Matrix December 2018 International Journal of Applied Mathematics and Computer Science 28(4):771-786 DOI:10.2478/amcs-2018-0059.
23. Coleman T., Saunderson J., and Wirth A. A local-search 2-approximation for 2-correlationclustering // LNCS. 2008. V. 5193. P. 308–319.
24. Donath W. E. and Hoffman A. J., lower bounds for the partitioning of graphss, IBM, 1973.
25. Duan L. A local-density based spatial clustering algorithm with noise / Duan L ., Xu L., Guo F., Lee J., Yan B. // Information Systems. 2007. № 32. P. 978–986.

26. Fortunato S. Community detection in graphs / Santo Fortunato // Complex Networks and Systems Lagrange Laboratory, ISI Foundation, Viale S. Severo 65, 10133, Torino, I, Italy.
27. Fortunato, S. Complex Systems Lagrange Laboratory (CNLL) / Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi, Institute for Scientific Interchange (ISI), Viale S. Severo 65, 10133, Torino, Italy Received 30 May 2008; revised manuscript received 11 September 2008; published 24 October 2008 // Benchmark graphs for testing community detection algorithms
28. Fiedler, M. algebraic connectivity of graphs, // Czech. Math. Journal, 23 (1973).
29. Jain A., Murty M., Flynn P. Data clustering: A review // ACM Computing Surveys. — 1999. — Vol. 31, no. 3. — Pp. 264–323.
30. <http://snap.stanford.edu>
31. <http://www-personal.umich.edu/~mejn/netdata/>
32. vlado.fmf.uni-lj.si/pub/networks/data/bio/Yeast/Yeast.htm
33. Il'ev, V. P. Approximation algorithms for graph approximation problems, / Il'ev, V. P., Il'eva, S. D., Navrotskaya, A. A. // Diskretn. Anal. Issled. Oper., 18, No. 1, 41–60, 2011. Translated in J. Appl. Ind. Math., 5, No. 4, 569–581, 2011.
34. Newman. M. E. J. Modularity and community structure in networks. 2006.
35. Newman, M. E. J. Community structure in social and biological networks. // M. Girvan, M. E. J. Newman, 2002.
36. Ramathilagam S., Huang Y.-M. Extended Gaussian Kernel Version of Fuzzy c-Means in the Problem of Data Analyzing // Expert Systems with Applications. 2011. Vol. 38, No. 4. P. 3793–3805.
37. Page, L. The Anatomy of a LargeScale Hypertextual Web Search Engine /L. Page, S. Brin.
38. Satu Elisa Schaeffer Survey Graph clustering Laboratory for Theoretical Computer Science, Helsinki University of Technology TKK, P.O. Box 5400, FI-02015 TKK, Finland.
39. Ulrike von Luxburg. A tutorial on spectral clustering. Statistics and Computing, 17(4):395 NAK 416, 2007.

Федеральное государственное автономное
образовательное учреждение высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Институт математики и фундаментальной информатики
Кафедра высшей и прикладной математики

УТВЕРЖДАЮ

Заведующий кафедрой

 / С.Г. Мысливец

« 17 » июня 2021 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

ИССЛЕДОВАНИЕ ЗАДАЧ КЛАСТЕРИЗАЦИИ ГРАФА

Направление 01.04.02 Прикладная математика и информатика

Магистерская программа 01.04.02.06 Прикладная математика и информатика

в гуманитарных и социально-экономических науках

Научный руководитель

кандидат физико-математических наук,

доцент

 / Д.В. Семенова
17 июня 2021

Выпускник

 / М.К. Кумарбекова
17 июня 2021

Красноярск 2021