

Министерство науки и высшего образования РФ
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

« ____ » _____ 2024 г.

БАКАЛАВРСКАЯ РАБОТА

Направление 01.03.02 Прикладная математика и информатика

ПРИБЛИЖЁННЫЕ АЛГОРИТМЫ ДЛЯ ЗАДАЧ КЛАСТЕРИЗАЦИИ НА ГРАФАХ

Руководитель	доцент, кандидат физико- математических наук	Д. В. Семенова
Выпускник		В. И. Чуринова
Нормоконтролер		Т. Н. Шипина

Красноярск 2024

РЕФЕРАТ

Выпускная квалификационная работа (дипломная работа) по теме «Приближённые алгоритмы для задач кластеризации на графах» содержит 25 страниц, на которых 12 рисунков и 3 таблицы. В работе использовано 9 источников литературы.

ГРАФЫ, КЛАСТЕРИЗАЦИЯ, АЛГОРИТМЫ.

Основная цель работы – исследование задач кластеризации графов и анализ методов их решения. Для достижения этой цели были поставлены и решены следующие задачи:

1. Провести обзор литературы и изучить задачи кластеризации графов.
2. Изучить существующие формулировки задач кластеризации графов и алгоритмы их решения.
3. Провести вычислительные эксперименты.

В бакалаврской работе достигнуты следующие результаты:

- Рассмотрены различные подходы к решению задач кластеризации графов.
- Изучены задачи GC , GC_k и $GC_{\leq k}$ кластеризации графов.
- Исследованы шесть алгоритмов кластеризации графов: BBC , CSW , A_1 , A_2 , A_3 с приведением примеров их работы.
- Разработаны программы, реализующие указанные алгоритмы и проведены вычислительные эксперименты с использованием библиотек Python.

СОДЕРЖАНИЕ

Введение	3
1 Постановка задачи кластеризации графа	5
1.1 Основные определения и обозначения	6
1.2 Постановка задач кластеризации графа	7
1.3 Выводы по первой главе	8
2 Приближенные алгоритмы решения задач $GC_{\leq k}$	9
2.1 Приближённые алгоритмы решения задачи $GC_{\leq 2}$	10
2.2 Приближённые алгоритмы решения задач $GC_{\leq 3}$	14
2.3 Выводы по второй главе	17
3 Вычислительные эксперименты	18
3.1 Результаты сравнения алгоритмов для задач $GC_{\leq 2}$	18
3.2 Результаты сравнения алгоритмов для задач $GC_{\leq 3}$	20
3.3 Выводы по третьей главе	22
Заключение	24
Список использованных источников	25

ВВЕДЕНИЕ

Актуальность работы. Графовые структуры широко используются для моделирования и анализа сложных систем в различных областях:

- информационные технологии (моделирование сетей компьютеров, интернета, социальных сетей, анализ данных),
- транспорт и логистика (оптимизация транспортных сетей, маршрутов доставки, планирование графиков и маршрутов воздушного, морского и наземного транспорта),
- биология и химия (анализ различных биологических структур, моделирование молекулярных реакций),
- социология и психология (изучение взаимосвязей между людьми, анализ влияния и распространения информации в обществе),
- финансы и экономика (анализ рынков, взаимосвязей между компаниями, портфельного управления и выявления финансовых рисков),
- электроника и коммуникации (проектирование цифровых схем, сетей связи, сетей передачи данных).

Обработка графов с миллионами вершин и рёбер делает использование точных алгоритмов практически невозможным и требует разработки эффективных приближённых алгоритмов, способных оперативно находить решения с достаточным качеством. Это делает тему исследования актуальной и практически значимой. Существующие алгоритмы кластеризации и анализа графов имеют свои ограничения. Работа направлена на исследование уже существующих подходов и методов, которые могут улучшить качество и эффективность анализа.

Одной из ключевых задач при работе с графами является кластеризация, которая помогает выявлять сообщества или группы взаимосвязанных вершин и лучше понимать структуру сетей.

Задачи кластеризации на графах привлекают значительное внимание исследователей из-за их широкой применимости в различных областях. Ильев В. П. и его коллеги внесли существенный вклад в разработку приближённых алгоритмов для кластеризации графов [1–4]: исследования в работах [1, 2] показывают эффективность приближённых методов для решения NP-трудных

задач на больших графах. В [4] рассмотрены проблемы кластеризации графов с ограничениями на размеры кластеров, выявлены основные сложности и предложены решения. В книге о кластерном анализе [5] предложены фундаментальные методы и подходы, используемые в задачах кластеризации, что имеет важное практическое значение, а монография [6] – это теоретическая база для понимания задач кластеризации.

Таким образом, существует широкий спектр методов и подходов к решению задач кластеризации на графах, включая как точные, так и приближённые алгоритмы. Эти работы формируют основу для дальнейших исследований и разработок в данной области.

Цели и задачи. Основная цель работы – исследование задач кластеризации графов и анализ методов их решения. Для достижения этой цели необходимо выполнить следующие задачи:

1. Провести обзор литературы и изучить задачи кластеризации графов.
2. Изучить существующие формулировки задач кластеризации графов и алгоритмы их решения.
3. Провести вычислительные эксперименты.

Структура и объём бакалаврской работы. Общий объём диссертационной работы составляет 25 страниц, на которых содержится 12 рисунков и 3 таблицы. Использовано 9 источников литературы.

ЗАКЛЮЧЕНИЕ

Основная цель данной бакалаврской работы состоит в изложении основных концепций традиционных алгоритмов кластеризации, а также в анализе их преимуществ и недостатков. Учитывая разнообразие методов и пересечение областей исследований, представление исчерпывающего списка всех существующих алгоритмов кластеризации является затруднительным. Проведенный анализ позволяет эффективно подбирать соответствующую группу алгоритмов в зависимости от конкретной технической задачи.

В работе достигнуты следующие результаты:

- Рассмотрены различные подходы к решению задач кластеризации графов.
- Изучены задачи GC , GC_k и $GC_{\leq k}$ кластеризации графов.
- Исследованы шесть алгоритмов кластеризации графов: BBC , CSW , A_1 , A_2 , A_3 с приведением примеров их работы.
- Разработаны программы, реализующие указанные алгоритмы и проведены вычислительные эксперименты с использованием библиотек Python.

Основные положения работы и отдельные вопросы были выставлены на обсуждение на XIX Международной научной конференции студентов, аспирантов и молодых учёных «Перспективны Свободный - 2023» (Красноярск, Россия, 2023), а так же на заседаниях научного семинара кафедры высшей и прикладной математики Института математики и фундаментальной информатики Сибирского федерального университета в 2023-2024 годах.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Ильев, В. П. Алгоритмы приближённого решения одной задачи кластеризации графа / В. П. Ильев, С. Д. Ильева, Моршинин А. В. // ПДМ. – 2019. – № 45. – С. 64–77.
2. Ильев, В. П. and Ильева, С. Д. and Моршинин, А. В. 2-Приближённые алгоритмы для двух задач кластеризации на графах / Ильев, В. П. and Ильева, С. Д. and Моршинин, А. В. // Дискретн. анализ и исслед. опер. – 2020. – Т. 27, № 3. – С. 88–108.
3. Ильев, В. П. and Ильева, С. Д. and Навроцкая, А. А. Приближенные алгоритмы для задач аппроксимации графов, дискретный анализ и исследование операций / Ильев, В. П. and Ильева, С. Д. and Навроцкая, А. А. // Дискретн. анализ и исслед. опер. – 2011. – Т. 18, № 1. – С. 41–60.
4. Балджанова, Р. В. and Ильев, А. В. and Ильев, В. П. О сложности кластеризации графа в задаче с ограничениями на размеры кластеров / Балджанова, Р. В. and Ильев, А. В. and Ильев, В. П. // ПДМ. – 2023. – № 60. – С. 76–84.
5. Дюран, Б. and Оделл, П. Кластерный анализ / Дюран, Б. and Оделл, П. – Статистика, 1977. – С. 128.
6. Зыков, А. А. Основы теории графов / А. А. Зыков. – Вузовская книга, 2004. – С. 664.
7. Coleman, T. and Saunderson, J. and Wirth, A. A local-search 2-approximation for 2-correlation clustering / Coleman, T. and Saunderson, J. and Wirth, A. // LNCS. – Vol. 5193. – 2008. – P. 308–319.
8. Воронцов, К. В. Алгоритмы кластеризации и многомерного шкалирования. Курс лекций / К. В. Воронцов. – МГУ, 2007.
9. Деза, Е. and Деза, М. М. Энциклопедический словарь расстояний / Деза, Е. and Деза, М. М. – Наука, 2008. – С. 448.

Министерство науки и высшего образования РФ
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

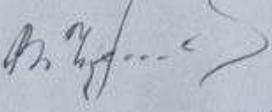
« 25 » 06 2024 г.

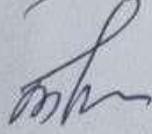
БАКАЛАВРСКАЯ РАБОТА

Направление 01.03.02 Прикладная математика и информатика

ПРИБЛИЖЁННЫЕ АЛГОРИТМЫ ДЛЯ ЗАДАЧ КЛАСТЕРИЗАЦИИ НА ГРАФАХ

Руководитель  доцент, кандидат физико- Д. В. Семенова
математических наук

Выпускник  В. И. Чуринова

Нормоконтролер  Т. Н. Шипина

25.06.24

Красноярск 2024