

УДК 004

https://doi.org/10.33619/2414-2948/91/55

КЛАСТЕРИЗАЦИЯ ГРАФОВ ДЛЯ ОБРАБОТКИ В GPU

©**Темников Д. О.**, ORCID: 0009-0005-0594-323X, Поволжский государственный университет телекоммуникаций и информатики, г. Самара, Россия, dimon.temnikoff@yandex.ru

©**Лиманова Н. И.**, ORCID: 0000-0003-2924-5602, SPIN-код: 9799-8380, д-р техн. наук, Поволжский государственный университет телекоммуникаций и информатики, г. Самара, Россия, nataliya.i.limanova@gmail.com

©**Козлов В. В.**, ORCID: 0000-0003-3735-9423, канд. техн. наук, Поволжский государственный университет телекоммуникаций и информатики, г. Самара, Россия, vco2005@mail.ru

CLUSTERING GRAPHS FOR GPU PROCESSING

©**Temnikov D.**, ORCID: 0009-0005-0594-323X, Povolzhskiy State University of Telecommunications and Informatics, Samara, Russia, dimon.temnikoff@yandex.ru

©**Limanova N.**, ORCID: 0000-0003-2924-5602, SPIN-code: 9799-8380, Dr. habil., Povolzhskiy State University of Telecommunications and Informatics, Samara, Russia, nataliya.i.limanova@gmail.com

©**Kozlov V.**, ORCID: 0000-0003-3735-9423, Ph.D., Povolzhskiy State University of Telecommunications and Informatics, Samara, Russia, vco2005@mail.ru

Аннотация. В данной статье рассмотрена многоуровневая кластеризация графов, используемая для обработки данных в GPU. Качество разбиения графа на разделы может оказать значительное влияние на общую производительность программного обеспечения. Поэтому очень важно быстро найти корректное разбиение графа на подграфы. В статье описаны области применения кластеризации, особенности и типы кластеризации, а также сделаны выводы об актуальности использования кластеризации графов в современных сферах человеческой деятельности.

Abstract. This article discusses the multilevel clustering of graphs used for data processing in the GPU. The quality of graph partitioning can have a significant impact on the overall performance of the software. Therefore, it is very important to quickly find the correct partition of the graph into subgraphs. The article describes the application areas of clustering, features and types of clustering, and also draws conclusions about the relevance of using graph clustering in modern spheres of human activity.

Ключевые слова: граф, кластеризация, графический процессор, анализ, кластер.

Keywords: graph, clustering, GPU, analysis, cluster.

Графы — это математические структуры, используемые для моделирования многих типов взаимосвязей и процессов в физических, биологических, социальных и информационных системах. Они также используются при решении различных задач высокопроизводительных вычислений и анализа данных. Вычислительные требования, связанные с крупномасштабной обработкой графов для кибераналитики, геномики, анализа социальных сетей и других областей требуют мощной вычислительной производительности,

которую могут обеспечить только ускорители. С появлением CUDA 8 компания NVIDIA представила новую библиотеку графических алгоритмов nvGRAPH, ускоряемых графическим процессором. Его первый выпуск, nvGRAPH 1.0, поддерживает 3 ключевых графовых алгоритма: PageRank, поиск кратчайшего пути из одного источника и поиск в ширину из одного источника [1].

Многим приложениям необходимо разбивать графы на подграфы или находить кластеры внутри них. Например, разбиение графа может быть использовано при численном решении уравнений в частных производных для выполнения более эффективного разреженного матрично-векторного умножения, а кластеризация графа может быть использована для моделирования, идентификации сообществ в социальных сетях и для обеспечения кибербезопасности (Рисунок).

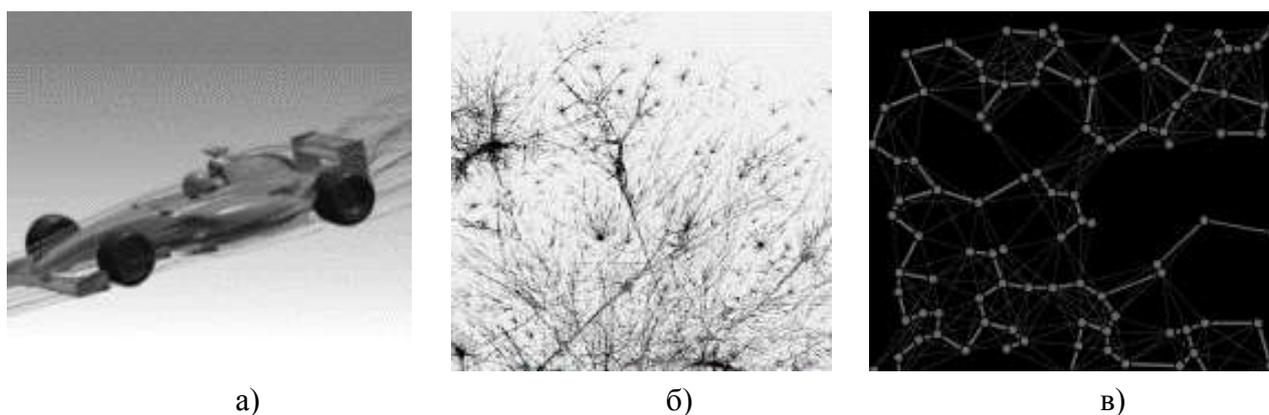


Рисунок. Применение разбиения графа на разделы: а) моделирование б) анализ социальных сетей в) кибербезопасность

Качество разбиения графа на разделы может оказать значительное влияние на общую производительность программного обеспечения. Поэтому очень важно не только быстро найти разбиение на подграфы, используя преимущества графических процессоров (схема разбиения спектрального графа на GPU работает в 7 раз быстрее, чем реализация на CPU), но и найти наилучшее возможное разбиение, что требует разработки новых алгоритмов.

Кроме того, разбиение графа на разделы и кластеризация направлены на поиск разбиения графа на подграфы на основе определенной метрики. В частности, разбиение спектрального графа и кластеризация основаны на спектре — собственных значениях и связанных с ними собственных векторах — матрицы Лапласа, соответствующей данному графу.

Параллельное разбиение графов является ключевым средством как для крупномасштабной аналитики графов, так и для высокопроизводительных научных вычислений. Разбиение графа — это задача создания непересекающихся наборов вершин в графе приблизительно одинакового размера при одновременном минимизации размера разреза, количества ребер, соединяющих вершины в разных наборах. Большинство программ и алгоритмов для разбиения графов на разделы используют многоуровневую эвристику. Многоуровневая кластеризация строит последовательность постепенно уменьшающихся графов на этапе укрупнения, находит решение проблемы (в данном случае секционирование) на наименьшем графе, а затем разматывает решение, чтобы найти решение для исходного графа. Этап разукрупнения также улучшает решение, используя информацию из каждого графа в последовательности в процессе, называемом уточнением. На высоком уровне

алгоритмы уточнения разбиения графа работают путем перемещения вершин для улучшения качества решения. Проблема уточнения разбиения графа хорошо изучена в контексте алгоритмов с общей памятью для многоядерных систем (<https://developer.nvidia.com/blog/fast-spectral-graph-partitioning-gpus/>).

Многоуровневая кластеризация является доминирующей стратегией для высококачественного последовательного и параллельного разбиения графа. Уточнение секционирования является ключевым этапом многоуровневого разбиения графа.

Многоуровневая кластеризация широко используется в крупномасштабном анализе графов. Программное обеспечение включает разбиение графов, кластеризацию, рисование и изучение представлений. Семейство алгебраических многосеточных методов и методов многоуровневой декомпозиции предметной области в линейной алгебре тесно связаны с многоуровневыми методами анализа графов. В многоуровневом методе вместо решения задачи на большом графе мы строим иерархию графов, которые постепенно уменьшаются по сравнению с исходным графом и при этом сохраняют структуру исходного графа. Затем мы решаем задачу на наименьшем графе и проецируем или интерполируем решение на исходный граф, используя иерархию.

Большинство графических разделителей предназначены для процессоров CPU и не работают на графических процессорах GPU. В частности, доработку многоуровневого алгоритма трудно распараллелить на графических процессорах. Первый известный разделитель графических процессоров был разработан Фаггингером, Ауэром и Бисселингом [2]. Они разработали два алгоритма для графического процессора: один многоуровневый спектральный, другой - многоуровневое уточнение на основе жадного алгоритма. Их библиотека так и не была выпущена. Более поздний разделитель графического процессора реализовал многоуровневый алгоритм с алгоритмом уточнения, основанным на распространении меток.

Sphynx — это спектральный разделитель, который работает на несколько графических процессоров. Это не многоуровневый процесс. Хотя он работает довольно быстро на графических процессорах, качество вырезания значительно хуже (примерно в 50 раз), чем у Metis / ParMETIS на нерегулярных графах.

Подводя итог, хочется отметить, что тема разбиения графов является актуальным направлением во многих сферах деятельности, например математический анализ или моделирование различных процессов.

Разработка эффективных методов разделения графов позволит решать многочисленные задачи с высокой эффективностью, что в свою очередь сможет принести большую выгоду компаниям, использующим эти методы. Поэтому многими разработчиками, например, компаниями NVIDIA и AMD в данный момент ведется разработка более эффективных методов кластеризации графов.

Список литературы:

1. Головченко Е. Н. Обзор алгоритмов декомпозиции графов. М.: Препринты ИПМ им. М. В. Келдыша, 2020. 38 с.
2. Gilbert M. S., Madduri K., Boman E. G., Rajamanickam S. Jet: Multilevel Graph Partitioning on GPUs // arXiv preprint arXiv:2304.13194. 2023. <https://doi.org/10.48550/arXiv.2304.13194>

References:

1. Golovchenko, E. N. (2020). *Obzor algoritmov dekompozitsii grafov*. Moscow. (in Russian).
2. Gilbert, M. S., Madduri, K., Boman, E. G., & Rajamanickam, S. (2023). *Jet: Multilevel Graph Partitioning on GPUs*. *arXiv preprint arXiv:2304.13194*. <https://doi.org/10.48550/arXiv.2304.13194>

*Работа поступила
в редакцию 19.05.2023 г.*

*Принята к публикации
25.05.2023 г.*

Ссылка для цитирования:

Темников Д. О., Лиманова Н. И., Козлов В. В. Кластеризация графов для обработки в GPU // Бюллетень науки и практики. 2023. Т. 9. №6. С. 470-473. <https://doi.org/10.33619/2414-2948/91/55>

Cite as (APA):

Temnikov, D., Limanova, N., & Kozlov, V. (2023). Clustering Graphs for GPU Processing. *Bulletin of Science and Practice*, 9(6), 470-473. (in Russian). <https://doi.org/10.33619/2414-2948/91/55>