

ОТЗЫВ

**официального оппонента на диссертационную работу
Афанасьева Ильи Викторовича**

**«Исследование и разработка методов эффективной реализации
графовых алгоритмов для современных векторных архитектур»,**

**представленную на соискание ученой степени
кандидата физико-математических наук
по специальности 05.13.11 – «Математическое и программное
обеспечение вычислительных машин, комплексов и
компьютерных сетей»**

Актуальность темы

В современном мире графы являются важной моделью представления данных, описывающей взаимодействующие или взаимосвязанные объекты реального мира или информационного пространства. Такие модели возникают при анализе структуры социальных сетей, перекрестных ссылок в интернет, публикаций в электронных библиотеках, взаимодействий агентов в задачах информационной или финансовой безопасности, взаимосвязей географических объектов, в естественнонаучных задачах и в других прикладных областях. Отличительной особенностью таких графов является большой размер: графы могут содержать миллионы вершин и миллиарды дуг, а значит, требуются эффективные инструменты для их представления и хранения. При этом подобные графы, как правило, имеют нетривиальную, нерегулярную структуру, включая в себя вершины с существенно различающимся числом соседей. Помимо эффективного представления и хранения графовой информации, важным является эффективность алгоритмов анализа структуры графов большого размера, в том числе поисковые алгоритмы (например,

алгоритмы поиска кратчайшего пути), алгоритмы расчета свойств вершин и дуг (например, алгоритмы расчета мер центральности или «репутационных» метрик типа PageRank, HITS и других), алгоритмы расчета потоков и прочие. Отличительной особенностью большинства из этих алгоритмов является то, что они не производят сложных вычислений над множеством отдельных вершин или дуг графа, но в тоже время требуют большого числа обращений к памяти для того, чтобы «перебрать» необходимое подмножество дуг и вершин для решения поставленной задачи. Поэтому работа с памятью становится «бутылочным горлышком» для таких алгоритмов, особенно на системах массового параллелизма, поскольку многие из этих алгоритмов требуют обхода всего графа или значительной его части. Для реальных задач проблема становится настолько сложной, что для того, чтобы ее решить за разумное время, зачастую вместо точных алгоритмов анализа используют приближенные методы, модели на основе случайных графов и другие, что естественно ухудшает качество результатов анализа. В то же время, очевидно, что алгоритмы анализа графов указанного выше типа обладают потенциально большими возможностями по распараллеливанию, в том числе на популярных и относительно доступных архитектурах, на основе SIMD подхода, использующих технологию «быстрой памяти» стандарта HBM. Именно на решение этой задачи важной и актуальной задачи – на разработку методов создания эффективных алгоритмов анализа графовых данных на векторных архитектурах с быстрой памятью – направлена представленная диссертационная работа И.В. Афанасьева.

Содержание диссертационной работы

Текст представленной диссертации состоит из введения, пяти глав, заключения и списка литературы. Диссертация построена четко и логично, связи смысловой составляющей от главы к главе прослеживаются легко.

Диссертационную работу И.В. Афанасьев начинает с обзора в первой главе подходов и методов реализации графовых алгоритмов для векторных систем с быстрой памятью. Апелляция к вычислительным системам с быстрой памятью разумна и понятна – рассматриваемые алгоритмы принадлежат к классу *memory-bound*, а, значит, интенсивно работают с памятью, и вычислительные системы с быстрой памятью и высокой пропускной способностью, очевидно, будут иметь более высокие показатели эффективности. Понятна и ориентация на векторные системы – все компьютеры с быстрой памятью обладают теми или иными элементами векторной обработки. Описывается набор из девяти графовых задач, алгоритмы решения которых автор рассматривает в данной работе: поиск кратчайших путей, поиск в ширину, поиск компонент сильной связности, ранжирование страниц и другие. Рассматривается вопрос и о том, что будет входными данными для будущих реализаций, для чего автор формирует хорошую базу реальных графов разного размера с разными свойствами.

Вторая глава работы посвящена анализу свойств целевых вычислительных систем, и основное внимание уделяется компьютерам с быстрой памятью: NEC SX-Aurora TSUBASA (производитель японская компания NEC), NVIDIA P100 (Pascal), NVIDIA V100 (Volta) и Intel KNL. В каждой архитектуре выделяются элементы векторной обработки, и, в конечном итоге, каждая архитектура анализируется на предмет того, что нужно знать и уметь, чтобы для каждой из них

создавать эффективные приложения. Немаловажно и то, что в результате анализа выделяются три ключевых свойства архитектур (массивный параллелизм, необходимость поддержки локальности данных и векторная обработка), которые потом проходят красной нитью и через анализ свойств графовых алгоритмов, и через проектирование эффективных реализаций с помощью созданного автором графового фреймворка VGL.

Третья глава представляет основной интерес. Автор анализирует информационную структуру всего множества рассматриваемых графовых алгоритмов, на основе чего приходит к выводу, что можно выделить лишь четыре алгоритмические абстракции (или типовые алгоритмические структуры), с помощью которых можно выразить все эти алгоритмы данного типа. Это – очень интересный результат, который в дальнейшем автор усиливает демонстрацией того, что все эти алгоритмические абстракции обладают указанными выше тремя ключевыми свойствами. Таким образом, если найти способ эффективной реализации указанных абстракций, то это и составит основу метода для создания эффективных реализаций графовых алгоритмов на выбранных архитектурах.

На этом пути есть очевидное препятствие – структуры данных, отсутствующие в алгоритмах, но определяющие эффективность реализаций на современных компьютерных системах, и именно структуры данных должны адекватно отражать и локальность, и векторность. Это и стало следующим шагом – автор предлагает две абстракции структур данных, которые этими свойствами как раз и обладают, что детально описано в следующей, четвертой главе диссертационной работы. В данной главе излагаются принципы

построения фреймворка VGL для создания эффективных реализаций графовых алгоритмов через выше упомянутые алгоритмические абстракции и абстракции данных. Показываются типичные схемы использования графового фреймворка VGL, а также приводятся примеры реализаций графовых алгоритмов для векторного суперкомпьютера NEC TSUBASA.

Пятая глава завершает изложение работы. Автор провел большое число вычислительных экспериментов на различных платформах и алгоритмах, на уже существующих пакетах и на своей системе VGL, убедительно показав преимущество предложенного им метода, намного превосходящего существующие результаты.

В заключении формулируются основные результаты, полученные автором в ходе диссертационного исследования.

Научная новизна, обоснование, достоверность и практическая ценность полученных научных положений, выводов и рекомендаций

Основным результатом, обладающим научной новизной в диссертационной работе, можно считать предложенные автором четыре типовые алгоритмические структуры, с помощью которых удалось выразить все рассмотренные автором графовые алгоритмы. Данный результат также определяет исключительно высокую практическую значимость работы, поскольку эффективная программная реализация данных абстракций для выбранных архитектур позволяет автору успешно достичь поставленной в самом начале диссертационной работы цели – превосходства по выбранным критериям над существующими реализациями “в разы”. Это полностью подтверждает и высокую

практическую значимость выполненной работы, и эффективность предложенного в работе метода, и достоверность положений, на основе которых автором создан фреймворк VGL для реализации графовых алгоритмов. При этом сам фреймворк VGL поддерживает три важных качества в создаваемых с его помощью реализациях: производительность, переносимость и продуктивность, что является исключительно редким качеством.

Положения работы сформулированы четко, а проведенные автором эксперименты на различных платформах убедительно показывают существенные преимущества созданных реализаций над аналогами.

Основные результаты полно представлены в 11 научных работах, в том числе в 9 работах в рецензируемых научных изданиях, определенных Положением МГУ о присуждении научных степеней.

Замечания по представленной диссертационной работе

Несмотря на высокое качество представленной работы, необходимо отметить ряд замечаний к настоящей работе.

1. Не до конца ясно, как именно автор выбрал набор алгоритмов анализа графов и набора архитектур. Класс графических процессоров сегодня достаточно обширен, но в работе подробно рассматриваются только процессоры NVIDIA P100/V100. Было бы интересно понять, применимы ли результаты, полученные автором, к графическим процессорам, которые используются на обычных рабочих станциях, к новым процессорам A100, к графическим процессорам от AMD и так далее. Также хотелось бы иметь обоснование выбора рассмотренного в работе подмножества алгоритмов анализа графов. Например, почему выбран

PageRank, а не HITs, почему из более чем десятка алгоритмов расчета мер центральности вершин была выбрана только часть. Возможно, за этим стояла какая-то практическая задача, для реализации которой разрабатывался фреймворк. Указание на такую практическую задачу только добавило бы работе веса и с точки зрения прикладной значимости, и в части логической завершенности.

2. Из описания предложенных автором четырех алгоритмических абстракций представляется, что абстракция `compute` всегда является первым этапом абстракции `reduce`. Нельзя ли было полностью разделить эти абстракции? Также в работе не сформулированы ограничения на операции агрегации в `reduce`, а они, возможно, существуют.

3. В разработанном автором представлении графовых данных `VectCSR` явно фигурируют задаваемые пользователем пороги для определения уровня связности вершин: «низкий», «высокий» и т.д. Вершины в каждом из подмножеств имеют принципиально разные структуры хранения информации о связях и соседях. Представляется, что производительность алгоритмов анализа графа существенно зависит от этих порогов. Но в работе не дано общего сценария выбора этих порогов (кроме очевидных советов, что нужно их выбирать в зависимости от длины вектора в архитектуре), также экспериментально не исследовано, как пороги влияют на результат.

4. Присутствует некоторая небрежность и в логике изложения, и в оформлении работы:

- Главы работы не очень хорошо сбалансированы по объему, возможно, имело смысл объединить 4 и 5 главу в одну, посвятив ее реализации и применению фреймворка.

- Не очень понятно, зачем в качестве бенчмарка для сравнения взята библиотека Ligr, если она показывает настолько плохой результат по сравнению со всеми конкурентами.
- Начало раздела 4.1. во многом повторяет материалы раздела 1.4. Логичнее было бы дать ссылку на предыдущий раздел вместо повтора.
- В рисунке 2.1 есть ошибка в подписи, и для рисунков 2.1-2.4 имело бы смысл указать ссылку на источник, если они взяты из документаций на системы, а не рисовались самостоятельно.

Заключение

Высказанные замечания не являются принципиальными и никак не влияют на высокую оценку полученных результатов. Диссертационная работа И.В.Афанасьева «Исследование и разработка методов эффективной реализации графовых алгоритмов для современных векторных архитектур» представляет собой завершенное научное исследование. Работа удовлетворяет всем требованиям Положения МГУ к кандидатским диссертациям, ее содержание соответствует паспорту специальности 05.13.11 – “Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей”, а ее автор, Афанасьев Илья Викторович, заслуживает присуждения ученой степени кандидата физико-математических наук по рассматриваемой специальности.

Официальный оппонент:

Кандидат физико-математических наук,
доцент кафедры Интеллектуальных информационных технологий
факультета ВМК МГУ им. М.В. Ломоносова

Петровский Михаил Игоревич

“18” декабря 2020 года

Контактные данные:

Тел: +7(910) 470 7376

Email: michael@cs.msu.su

Специальность, по которой официальным оппонентом защищена
диссертация: 05.13.11 “Математическое и программное обеспечение
вычислительных машин, комплексов и компьютерных сетей”

Адрес места работы:

119991, Москва, ГСП-1, Ленинские горы, д.1, стр.52, ф-т. ВМК МГУ

Тел: +7 (499) 9391789

Email: michael@cs.msu.su

Подпись доцента каф. ИИТ ВМК МГУ им. М.В. Ломоносова

М.И.Петровского удостоверяю:



Подпись удостоверяю
Ведущий специалист по кадрам
Т.Г. Коваленко