Эффективные линейно-алгебраические алгоритмы поиска заданных подструктур в графах

И. В. Антонов

Научный руководитель:
д.ф.-м.н., проф.
Д. С. Малышев

30 июня 2022 г.

Цели и задачи

Цели и задачи:

- Изучение алгоритма Нешетрила и Поляка [1] поиска k-клик.
- Изучение стандарта GraphBLAS для имплементации линейно-алгебраических алгоритмов.
- Построение линейно-алгебраической версии алгоритма подсчета k-клик Нешетрила и Поляка, опираясь на работу Емелина и др. [2].
- Программная реализация параллельного линейно-алгебраического алгоритма подсчета k-клик с помощью стандарта GraphBLAS.

Алгоритм Нештерила и Поляка

Концепция алгоритма поиска k клик:

1. Число k раскладывается как $k = 3l + r$, где $0 \leq r \leq 2$.
2. Строится граф l-клик.
3. В графе l-клик находятся треугольники, соответствующие $3l$-кликам.
4. Если $r \neq 0$, то $3l$-клики дополняют до $3l + r$-клик путем добавления r вершин.

Графовые матрицы

Определение

Пусть выбран набор клик $\{Clique(i)\}_{i=1}^{p}$*. Матрицей инцидентности клик вершинам графа называется матрица* $\boldsymbol{C} \in \{0,1\}^{p \times n}$*, построенная по правилу:*

$$\boldsymbol{C}(i,j) = \begin{cases} 1, & j \in Clique(i) \\ 0, & j \notin Clique(i) \end{cases}$$

Определение

Клики $Clique(i)$ *и* $Clique(j)$ *называются смежными, если любая вершина из клики* $Clique(i)$ *смежна с любой вершиной из клики* $Clique(j)$*.*

Графовые матрицы

Определение

Пусть даны два набора клик $\{Clique_1(i)\}_{i=1}^{p_1}$, $\{Clique_2(j)\}_{j=1}^{p_2}$. *Матрицей смежности первого и второго наборов клик называется матрица* $\boldsymbol{M} \in \{0,1\}^{p_1 \times p_2}$, *которая задается по правилу:*

$$\boldsymbol{M}(i,j) = \begin{cases} 1, & Clique_1(i) \text{ и } Clique_2(j) \text{ смежны} \\ 0, & \text{иначе} \end{cases}$$

Будем обозначать через $\boldsymbol{M}_{l,r}$ - матрицу смежности всех l на r клик. В случае $l = 1$ и $r = 1$ данная матрица совпадает с $\boldsymbol{A}_G$ - матрицей смежности графа.

Вычисление матрицы смежности l-клик

- Дано: $\boldsymbol{C}_l$ – матрица инцидентности l-клик вершинам графа.
- Требуется: получить матрицу $\boldsymbol{M}_{l,l}$ смежности l-клик.
- Решение: вычислим матрицу $\boldsymbol{C}_l \boldsymbol{A}_G \boldsymbol{C}_l^T$, заменим в полученной матрице числа равные l^2 на 1, а остальные на 0.

Замечание

Отсутствует биекция между треугольниками в графе l-клик и $3l$-кликами в исходном графе.

Для установления биекции усилим определение смежности клик $\boldsymbol{C}_l(i,:)$ и $\boldsymbol{C}_l(j,:)$ условием[1]:

$$max(nnz(\boldsymbol{C}_l(i,:))) < min(nnz(\boldsymbol{C}_l(j,:)))$$

Данное преобразование легко осуществить в линейно-алгебраических терминах.

[1] $nnz(v)$ - множество индексов ненулевых элементов вектора v

Вычисление матрицы инцидентности $3l$-клик вершинам графа

- Дано: $\boldsymbol{C}_l$ – матрица инцидентности l-клик вершинам графа, $\boldsymbol{M}_{l,l}$ – матрица смежности l-клик.
- Требуется: найти матрицу $\boldsymbol{C}_{3l}$ инцидентности $3l$-клик вершинам графа.
- Решение: вычислим матрицу[2] $\boldsymbol{B} = \boldsymbol{M}_{l,l} * \boldsymbol{M}^2_{l,l}$. Если $\boldsymbol{B}(i,j) \neq 0$ и $[\boldsymbol{M}_{l,l}(i,:) * \boldsymbol{M}^T_{l,l}(j,:)](k) \neq 0$, то $\boldsymbol{C}_l(i,:) + \boldsymbol{C}_l(j,:) + \boldsymbol{C}_l(k,:)$ – $3l$-клика.

[2] $\boldsymbol{A} * \boldsymbol{B}$ – поэлементное умножение матриц $\boldsymbol{A}$ и $\boldsymbol{B}$

Вычисление матрицы инцидентности $l+1$-клик вершинам графа

- Дано: $\boldsymbol{C}_l$ – матрица инцидентности l-клик вершинам графа, $\boldsymbol{C}_1$ – матрица инцидентности 1-клик вершинам графа (единичная матрица)
- Требуется: найти матрицу $\boldsymbol{C}_{l+1}$ инцидентности $l+1$-клик вершинам графа.
- Решение: вычислим матрицу $\boldsymbol{B} = \boldsymbol{C}_l \boldsymbol{A}_G$. Если $\boldsymbol{B}(i,j) = l$, то $\boldsymbol{C}_l(i,:) + \boldsymbol{C}_1(j,:)$ – $l+1$-клика.

Подсчет $3l$, $3l+1$, $3l+2$ клик

- Дано: $\boldsymbol{C}_l$, $\boldsymbol{C}_{l+1}$ – матрицы инцидентности l и $l+1$ клик вершинам графа соответственно.
- Требуется: подсчитать количество $3l$, $3l+1$, $3l+2$ клик.
- Решение: вычислим матрицы $\boldsymbol{M}_{l,l}$, $\boldsymbol{M}_{l,l+1}$, $\boldsymbol{M}_{l+1,l}$, $\boldsymbol{M}_{l+1,l+1}$ смежности l и l клик, l и $l+1$ клик, $l+1$ и l клик, $l+1$ и $l+1$ клик соответственно. Подсчитаем количество клик из соотношений:
  1. Количество $3l$-клик:
  $$sum(\boldsymbol{M}_{l,l} * \boldsymbol{M}^2_{l,l})$$
  2. Количество $3l+1$-клик:
  $$sum(\boldsymbol{M}_{l+1,l} * (\boldsymbol{M}_{l+1,l}\boldsymbol{M}_{l,l}))$$
  3. Количество $3l+2$-клик:
  $$sum(\boldsymbol{M}_{l,l+1} * (\boldsymbol{M}_{l,l+1}\boldsymbol{M}_{l+1,l+1}))$$

Результаты вычислительных экспериментов

Алгоритм был реализован на языке **Python** с помощью библиотеки **python-graphblas**. Для проведения вычислительных экспериментов использовался высокопроизводительный вычислительный кластер **"cHARISMa"**(2 x Intel Xeon Gold 6152 2.1-3.7 ГГц (2*22 ядер), и 1536 ГБ оперативной памяти).

Результаты вычислительных экспериментов на графе **com-DBLP** [3] (317.080 вершин, 1.049.866 ребер).

k	время выполнения алгоритма (в секундах)	количество k-клик
3	0.16375	2224385
4	0.42499	16713192
5	1.91298	262663639
6	41.86686	4221802226
7	1461.2348	60913718813

При $k = 8$ алгоритму не хватило оперативной памяти для хранения матрицы смежности 3 на 3 клик.

Заключение

Вывод:

- Данный алгоритм хорошо показал себя на **com-DBLP** [3], однако присутствует проблема с нехваткой оперативной памяти при больших размерностях (подсчет k-клик при $k = 8$).

Дальнейшая перспектива:

- Изучение динамического алгоритма подсчета k-клик для решения проблем с нехваткой оперативной памяти.

Список использованной литературы

[1] Jaroslav Nesetril and Svatopluk Poljak.
On the complexity of the subgraph problem.

[2] Maxim Emelin, Ilya Khlystov, Dmitriy Malyshev, and Olga Razvenskaya.
On linear algebraic algorithms for the subgraph matching problem and its variants.

[3] Jure Leskovec and Andrej Krevl.
SNAP Datasets: Stanford large network dataset collection.
http://snap.stanford.edu/data, June 2014.

[4] Jeremy Kepner.
Graphblas mathematics - provisional release 1.0 -.

Программная реализация

Спасибо за внимание!