

Теория графов. Презентация 2

Ефремов Алексей, Сичкар Георгий, Кононова Юлия

В предыдущих сериях



Реализации модели Google pregel:

- Pregel+
- Apache Giraph
- Apache Spark

Алгоритмы:

- MS-Parent BFS
- Boruvka

Характеристики вычислительной машины



- **CPU:** Intel i7-10510U
 - 4 ядра
 - hyper-threading есть (**отключили через BIOS**)
 - все ядра равнозначны
- **RAM:** 48 Gb
 - без файла подкачки
- **OS:** Ubuntu MATE 22.04.5 LTS x86_64



Дополнительные инструменты

Apache Spark 3.5.0:

- hadoop 3.3.4
- scala 2.12.18
- java 11

Apache Giraph 1.3-SNAPSHOT:

- java8
- hadoop 2.10.2

PregeL+ (commit: t620ec13):

- java 11
- hadoop 2.10.2
- g++ 11.4.0
- mpich 4.1.2



Таблица: **MS-Parent BFS**

SNAP	Число вершин	Число ребер
soc-Pokec	1 632 803	30 622 564
twitch_gamers	168 114	6 797 557
ego-Twitter	81 306	1 768 149

Таблица: **Boruvka**

DIMACS	Число вершин	Число ребер
CAL	1 890 815	4 657 742
NE	1 524 453	3 897 636
NW	1 207 945	2 840 208

Эксперимент I



Какая реализация Pregel демонстрирует наилучший результат по скорости для «**алгоритма**»?

Ход эксперимента:

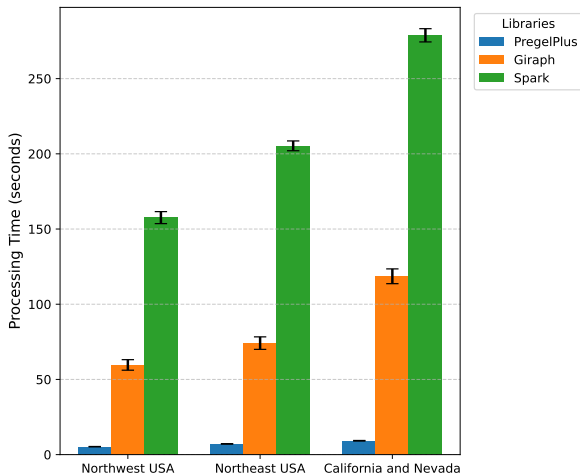
- Запустить алгоритм 20 раз
- Составить информацию о среднем значении и доверительных интервалах
- Визуализировать результаты

алгоритм: MS-Parent BFS; Boruvka



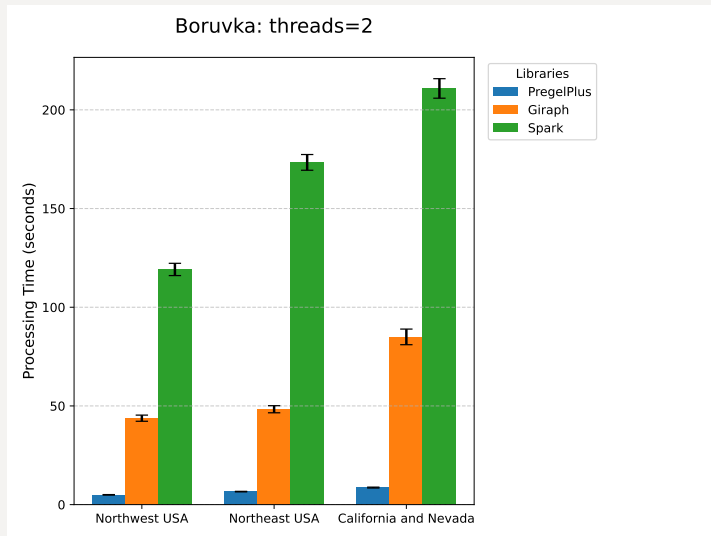
Boruvka (thread1)

Boruvka: threads=1





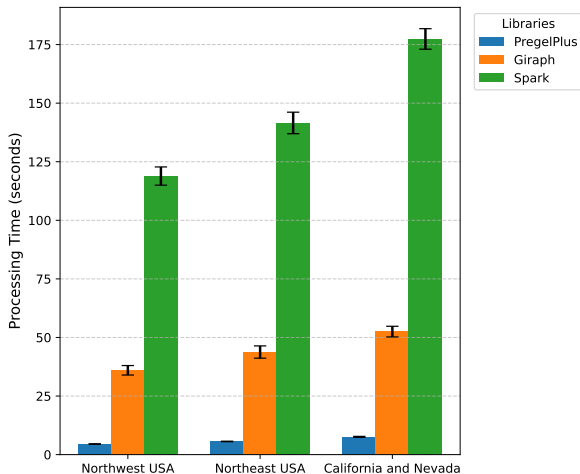
Boruvka (thread2)





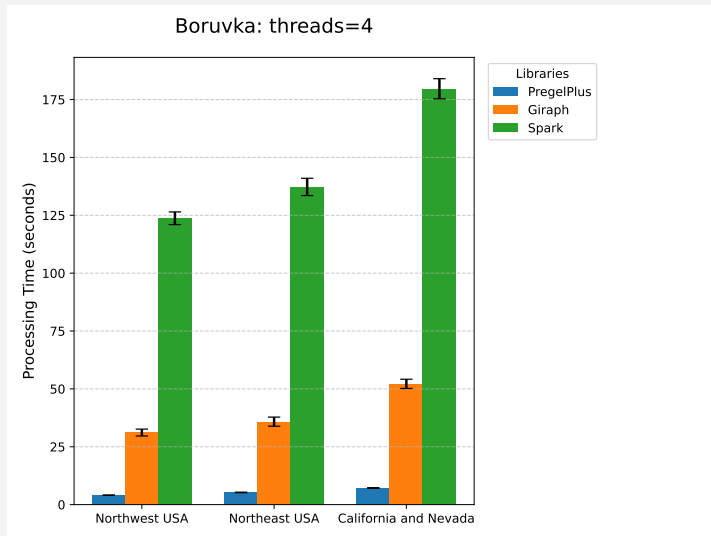
Boruvka (thread3)

Boruvka: threads=3



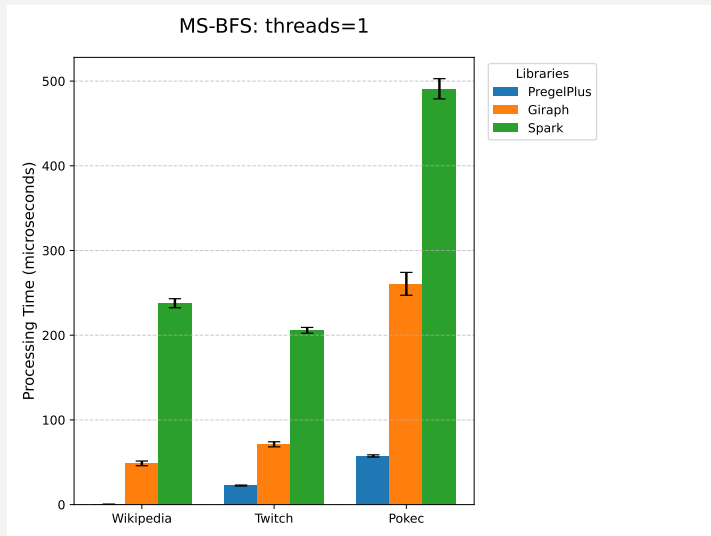


Boruvka (thread4)





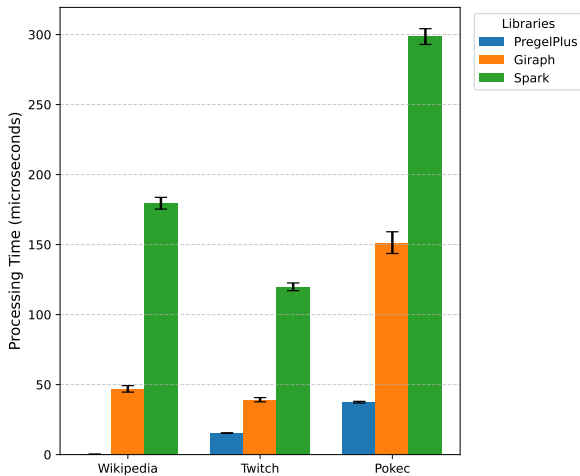
MS-Bfs (thread1)





MS-Bfs (thread2)

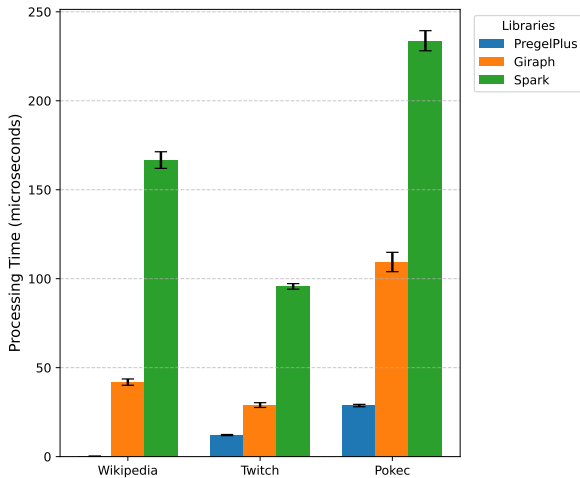
MS-BFS: threads=2





MS-Bfs (thread3)

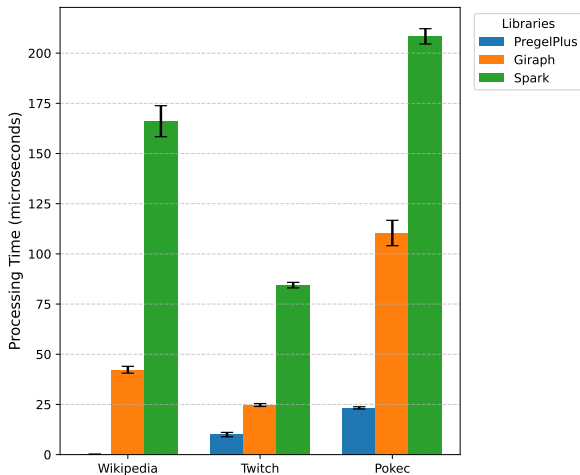
MS-BFS: threads=3





MS-Bfs (thread4)

MS-BFS: threads=4



Результаты (вывод)



Вывод:

Prege^l+ по скорости превосходит Apache Giraph и Apache Spark, в то время как Apache Giraph быстрее Apache Spark.

Эксперимент II

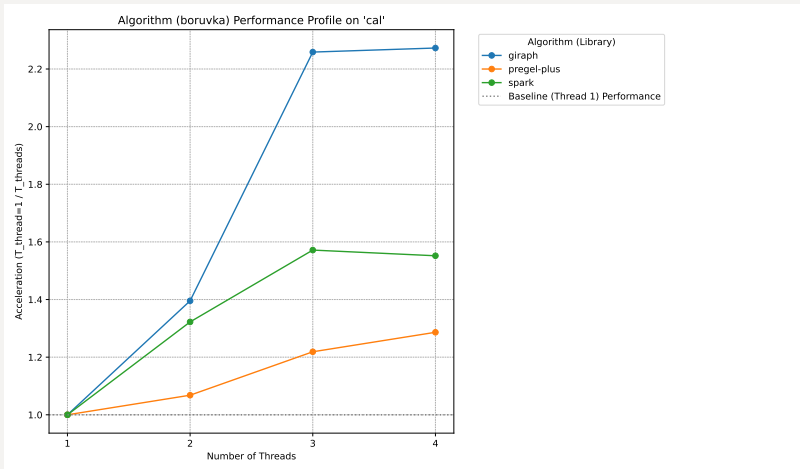


Насколько отличается ускорение «**алгоритма**» при увеличении числа выделенных потоков для различных реализаций Pregel?

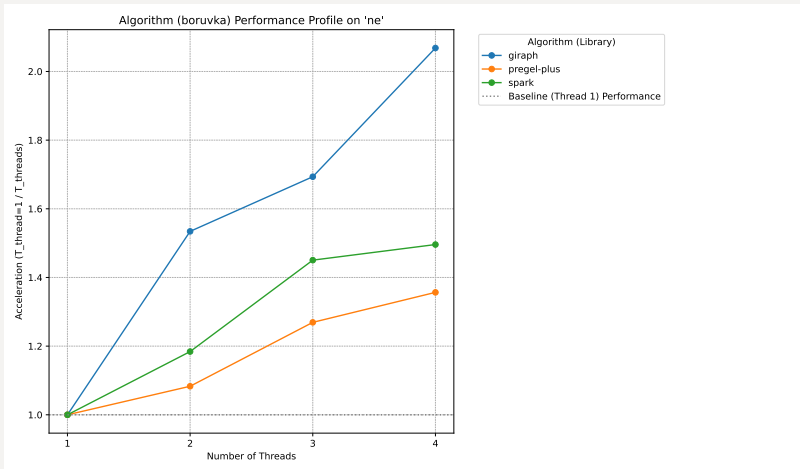
Ход эксперимента:

- Запустить алгоритм несколько раз
- Составить информацию о среднем значении и доверительных интервалах
- Визуализировать результаты

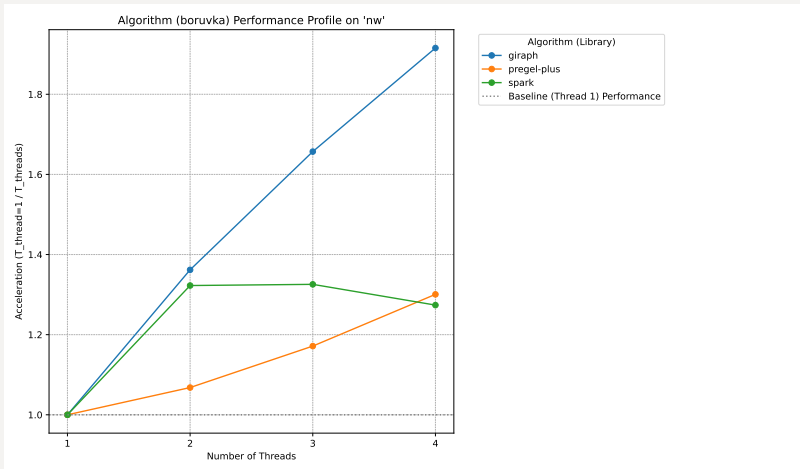
Boruvka (cal)



Boruvka (ne)

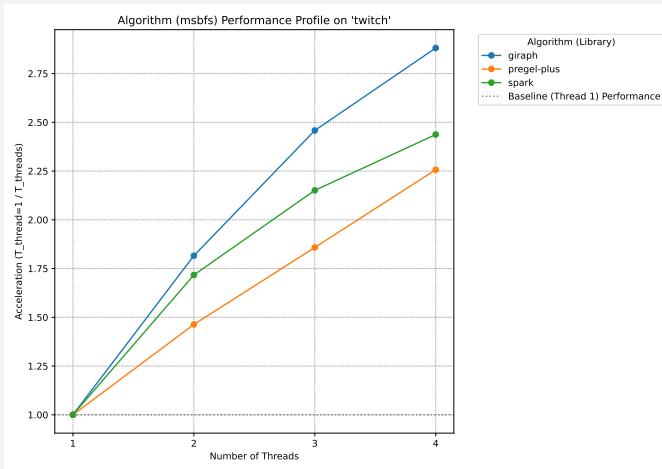


Boruvka (nw)

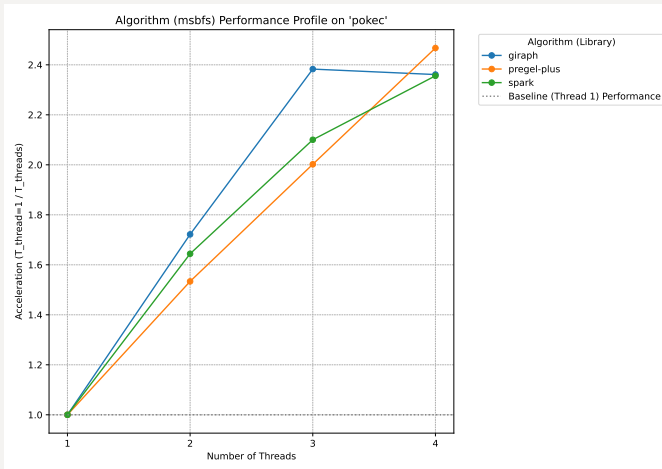




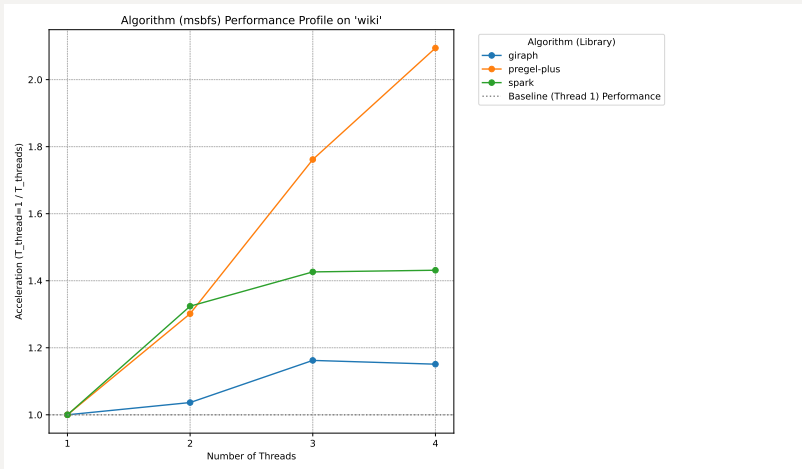
MS-Bfs (twitch)



MS-Bfs (pokec)



MS-Bfs (wiki)





Результаты (вывод)

Вывод:

Практически на всех графах **Apache Giraph** имеет наибольшее ускорение при увеличении числа потоков. На втором месте **Apache Spark**. Самый последний **Prege1+**.

исключение: wiki *MS-Parent BFS* из-за сильной связности графа

Корректность реализаций алгоритмов

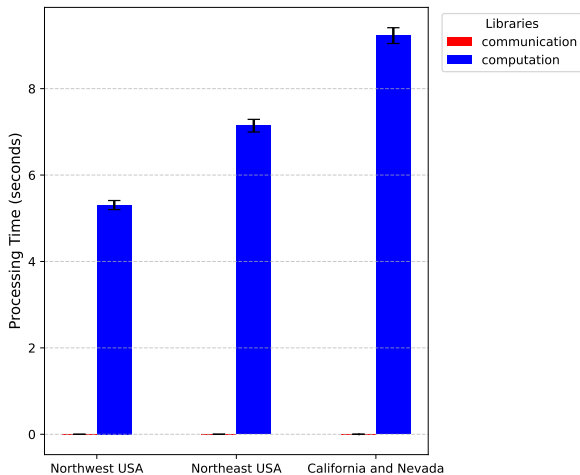


- **MS-Parent BFS**
- Перевод в единый формат:
- `id_i ipId_i ..`
- `id_j jpId_i ..`
- ...
- **Borůvka**
- Проверка суммы ребер MST на равенство
- Проверка краевых случаев на малых графах



MS-Bfs (thread1)

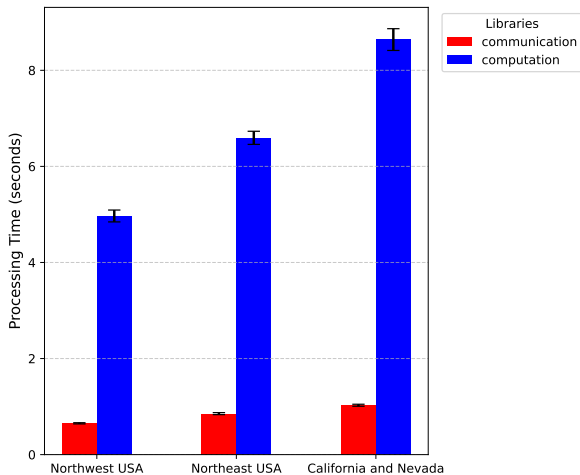
Boruvka (pregel-plus): threads=1





MS-Bfs (thread2)

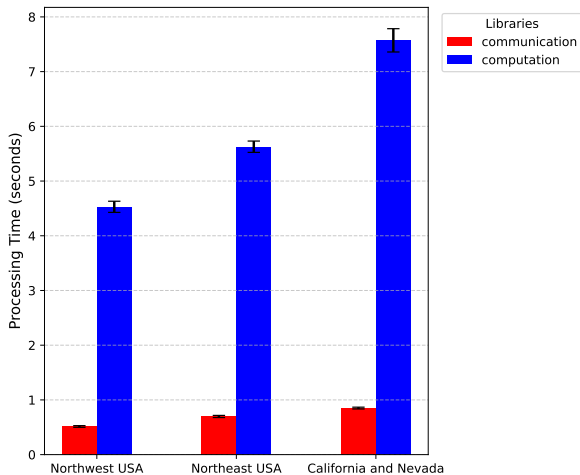
Boruvka (pregel-plus): threads=2





MS-Bfs (thread3)

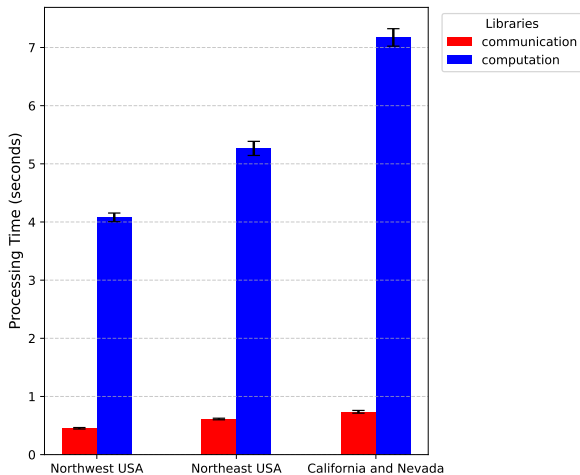
Boruvka (pregel-plus): threads=3





MS-Bfs (thread4)

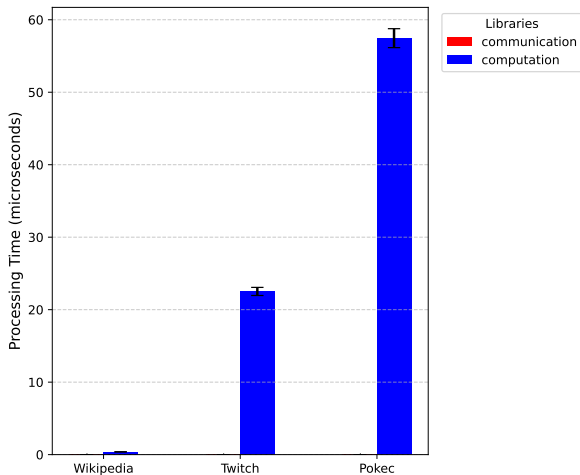
Boruvka (pregel-plus): threads=4





Boruvka (thread1)

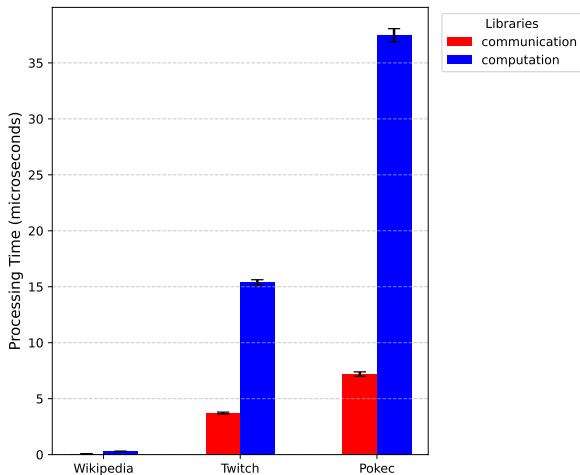
MS-BFS (pregel-plus): threads=1





Boruvka (thread2)

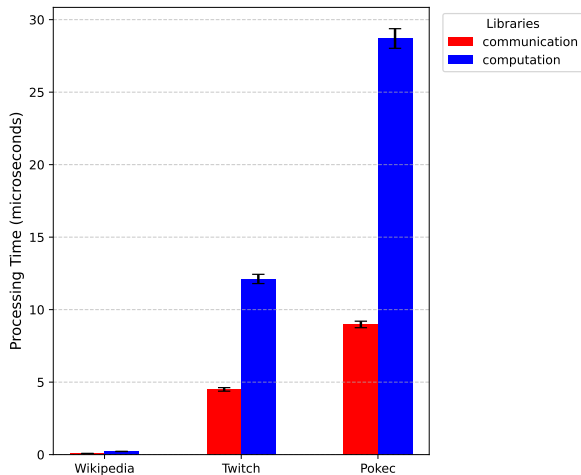
MS-BFS (pregel-plus): threads=2





Boruvka (thread3)

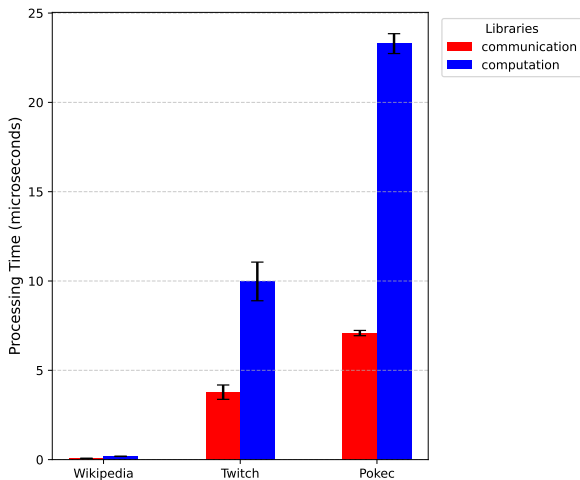
MS-BFS (pregel-plus): threads=3





Boruvka (thread4)

MS-BFS (pregel-plus): threads=4



boruvka: giraph



ThreadN	Time	acceleration
1	118 585 \pm 4 913	1,00
2	84 976 \pm 3 956	1,40
3	52 500 \pm 2 275	2,26
4	52 175 \pm 1 994	2,27

boruvka: pregel-plus



ThreadN	Time	acceleration
1	9 227 \pm 182	1,00
2	8 639 \pm 226	1,07
3	7 571 \pm 212	1,22
4	7 173 \pm 150	1,29

boruvka: spark



ThreadN	Time	acceleration
1	278 809 \pm 4 400	1,00
2	210 846 \pm 4 944	1,32
3	177 404 \pm 4 376	1,57
4	179 660 \pm 4 353	1,55

msbfs: giraph



ThreadN	Time	acceleration
1	260 663 \pm 13 490	1,00
2	151 397 \pm 7 771	1,72
3	109 389 \pm 5 436	2,38
4	110 399 \pm 6 325	2,36

msbfs: pregel-plus



ThreadN	Time	acceleration
1	57 460 \pm 1 306	1,00
2	37 466 \pm 589	1,53
3	28 700 \pm 676	2,00
4	23 289 \pm 555	2,47

msbfs: spark



ThreadN	Time	acceleration
1	490 908 \pm 11 933	1,00
2	298 581 \pm 5 627	1,64
3	233 738 \pm 5 649	2,10
4	208 363 \pm 3 818	2,36

Наши специалисты



Леша.png



Юля.webm



Гоша.jpg