TriX: Triangle Counting at Extreme Scale

Yang Hu Pradeep Kumar Guy Swope H. Howie Huang

HPEC'17
September 13, 2017

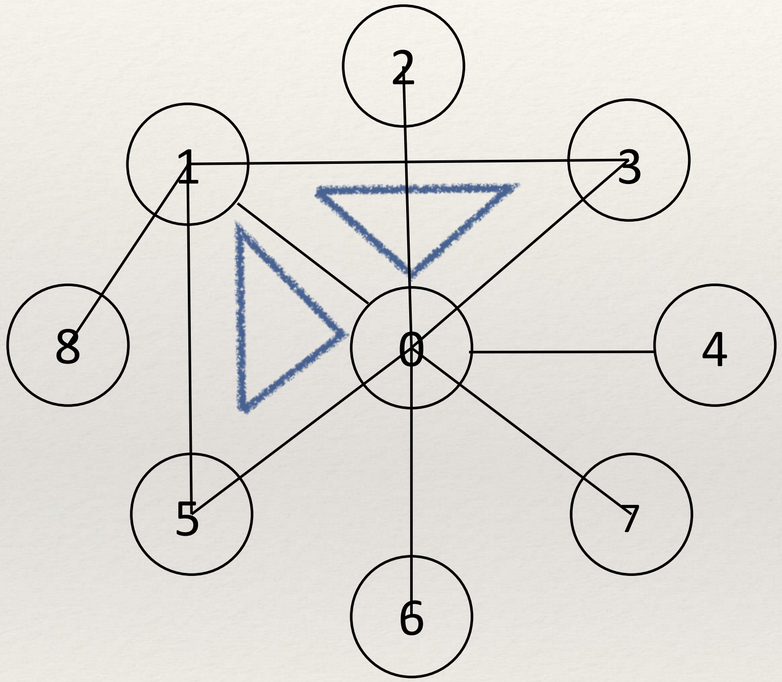
THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Scaling to Big Graphs

- Challenge #1: High computational cost of triangle counting
  - Fast GPU based triangle counting
- Challenge #2: Extremely large graph cannot fit in memory
  - External memory algorithm using 2-D partitioning

GW

Background: Triangle Counting Algorithm



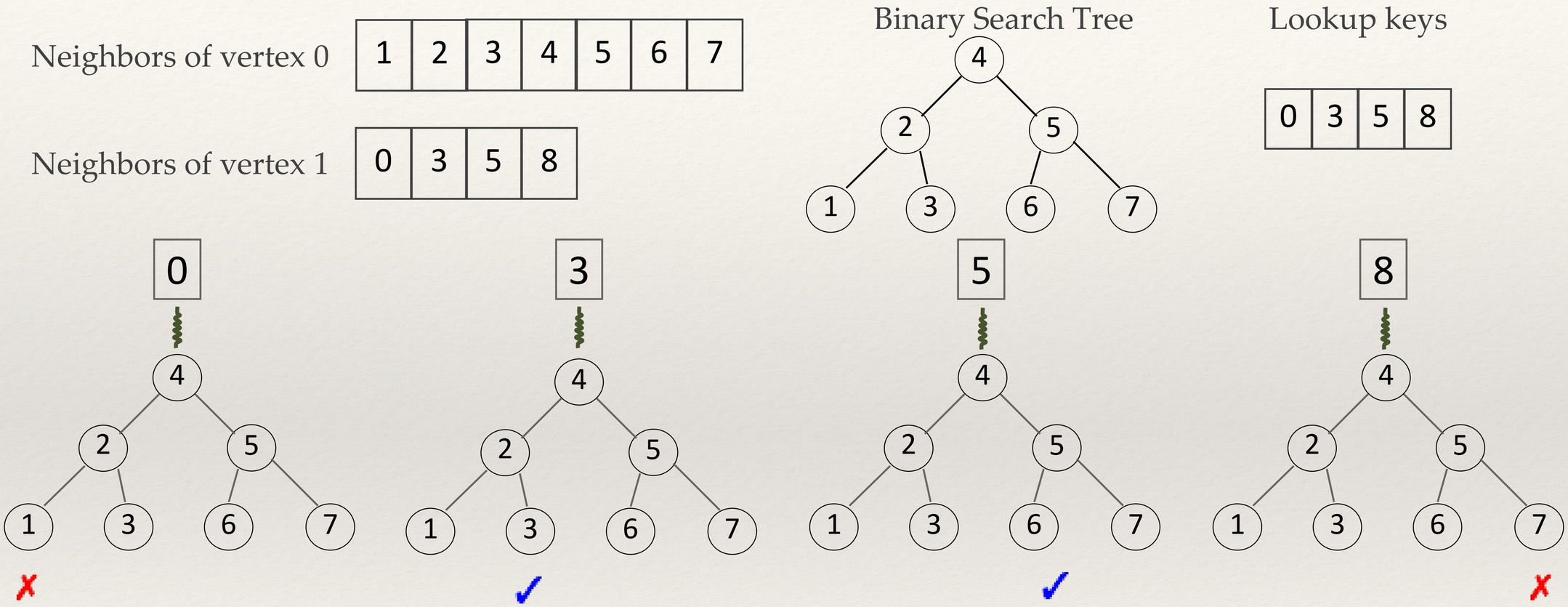
- Intersection on each edge

Neighbors of vertex 0	1	2	3	4	5	6	7

Neighbors of vertex 1	0	3	5	8

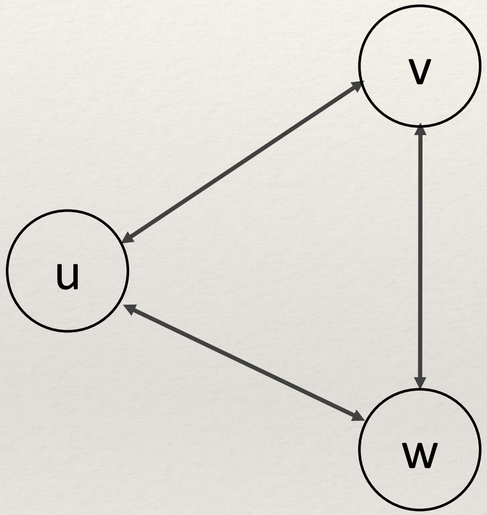
- Two Triangles (0,1,3) and (0,1,5)
  - Start from edge (0,1)
  - Compare the neighbors of both vertices
  - Find shared neighbors 3 and 5

Binary Search Based Triangle Counting

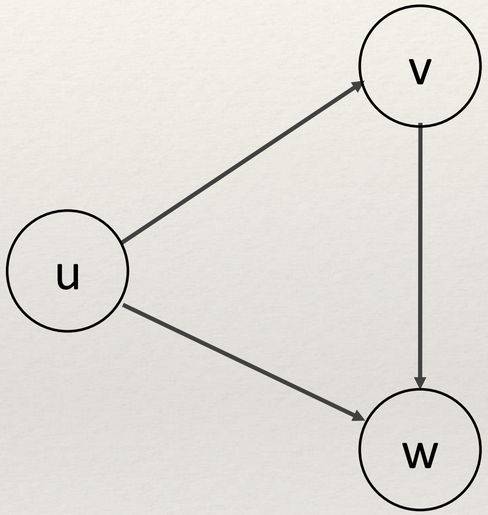


- Fine-grained parallelization on GPUs
- Better locality for faster memory access

Background: Orientation and Directed Triangle



- Orientation
  - Topology order
  - Rank-by-degree
- Directed triangle
  - Avoid redundant work



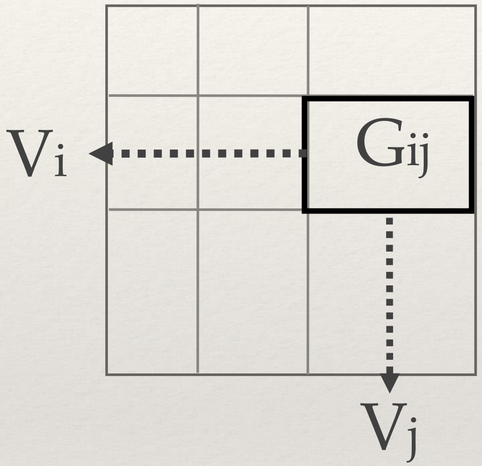
External Memory Triangle Counting

- Assumption: small internal memory size and unlimited external storage (disk) size
- To find a triangle, three edges have to be in memory at same time
  - Read everything from disk?
- Minimize external I/Os, while ensuring algorithm correctness

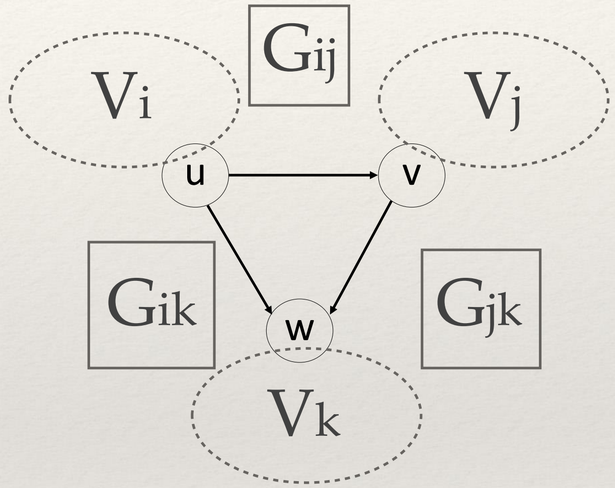
GW

Naive 2-D Partitioning

- 2-D partitioning



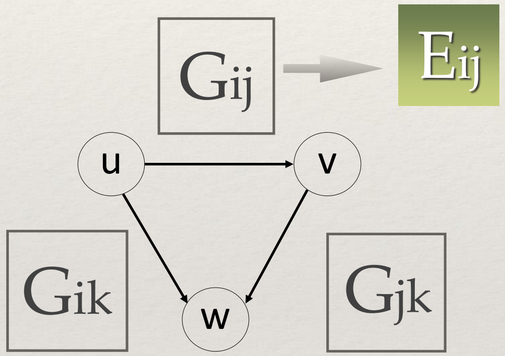
- Sub-task

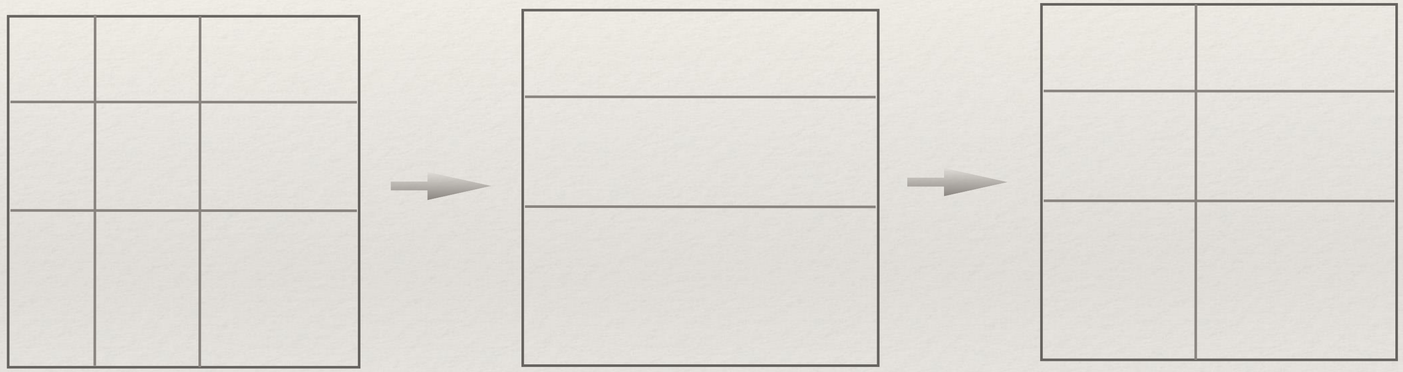


- Achieves optimal I/O complexity
- Support massive parallelism
- Partitions are not balanced!

2-D Managed Partition

- The edge contains first two vertices can be streamed.
- The 1-D partitions can be balanced simply.

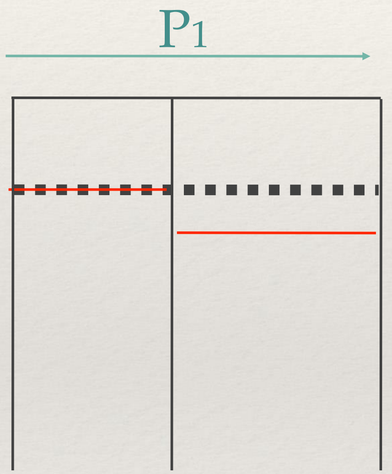




- Use edge list for the first edges and focus on balancing the CSR partitions.
- Partition the CSR columns and rows by different vertex ranges.

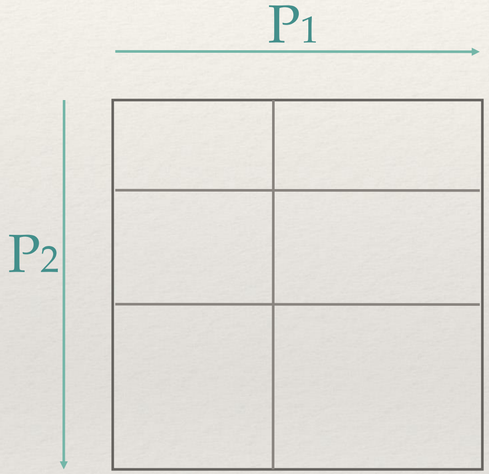
2-D Managed Partition

Balanced column-wise partition of CSR



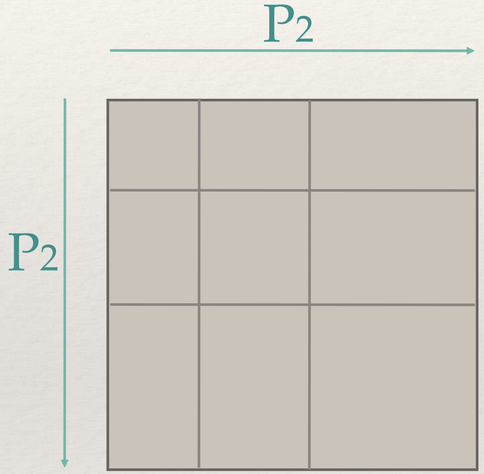
CSR

Row-wise partition to satisfy the memory bound

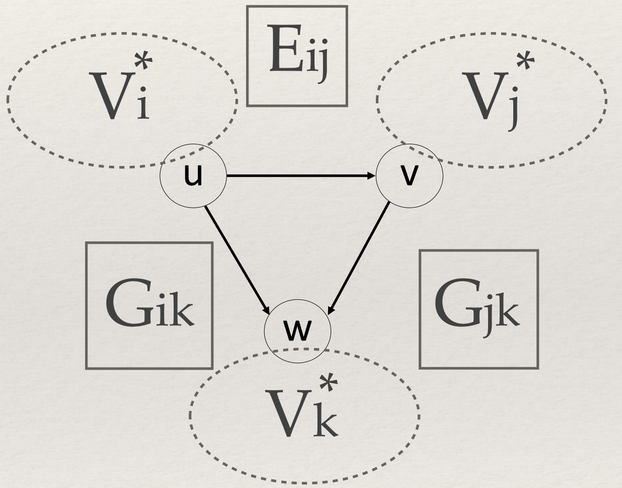


CSR

Corresponding edge partitioning



Edge List



Experiments

- Environments
  - A cluster: each node with dual Intel Xeon E-2620 2.0GHz 6-core processors, 128GB memory and NVIDIA K20 GPUs
  - A server with dual Intel Xeon E5-2683 CPU, 512GB memory and NVIDIA K40 GPUs
- Datasets
  - 13 graphs (real and synthetic), up to billion nodes and hundreds GB

Graph Datasets

Dataset	Description	|V|	|E|	Triangles
email-EuALL	Email network	265,215	364,481	267,313
soc-Epinions1	Who-trusts-whom network	75,880	405,740	1,624,481
flickerEdges	Image relationships	105,939	2,316,948	107,987,357
RoadNet	Road network of California	1,965,207	2,766,607	120,676
graph500-scale18-ef16	Synthetic	174,148	3,800,348	82,287,285
cit-Patents	Citation network US patents	3,774,769	16,518,947	7,515,023
Orkut	Social network	33,554,432	523,611,003	22,535,831,016
Kron-25-16	Synthetic	8,730,857	327,036,486	223,127,577
Friendster	Social network	41,652,230	1,202,513,046	34,824,916,864

TC Performance (Single GPU)

Dataset	Time (seconds)	MTEPS
email-EuALL	0.003009322	121.12
soc-Epinions1	0.004831075	84.08
flickerEdges	0.004825446	32.59
RoadNet	0.0196982	140.45
graph500-scale18-ef16	0.126783	29.98
cit-Patents	0.1280756	128.98
Orkut	10.65	30.7
Kron-25-16	72	7.27

GW

TC Performance (32 GPUs)

Dataset	Time (seconds)	MTEPS
Friendster	17.12	70.25
Twitter MPI	43.51	41.64
Gsh-2015	2029.03	16.4
Kron-30-16	736.43	23.11
Kron-31-16	2079.33	16.4

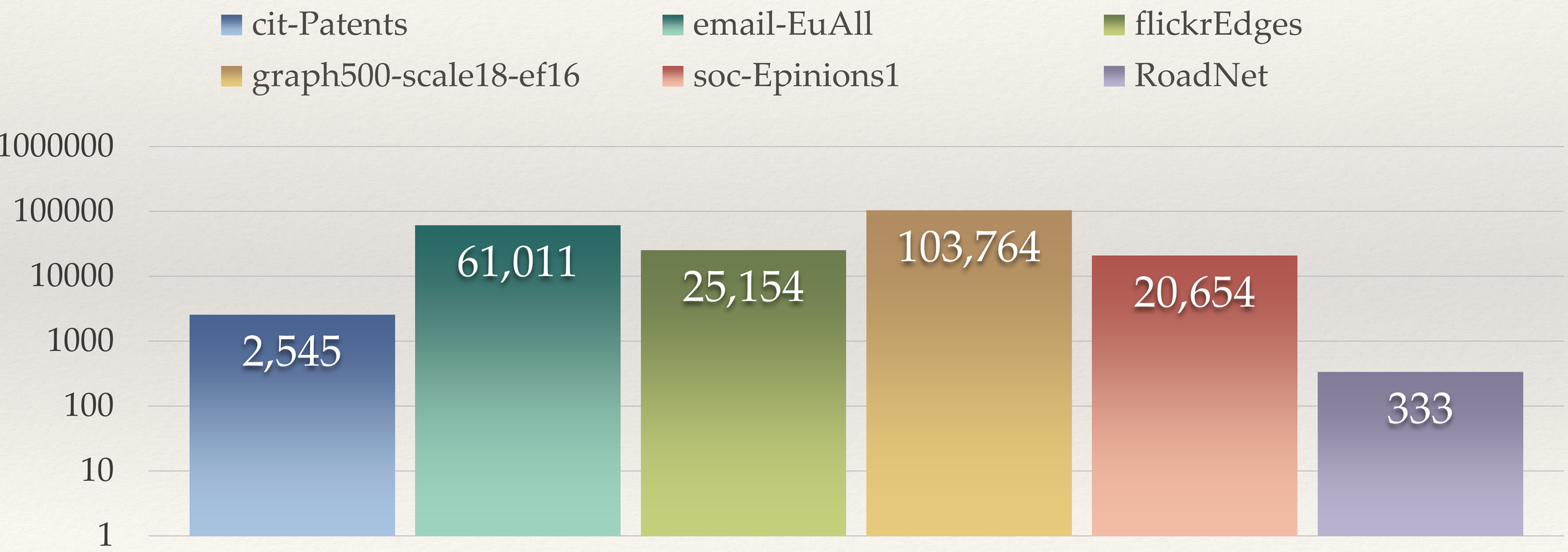
GW

Power Consumption

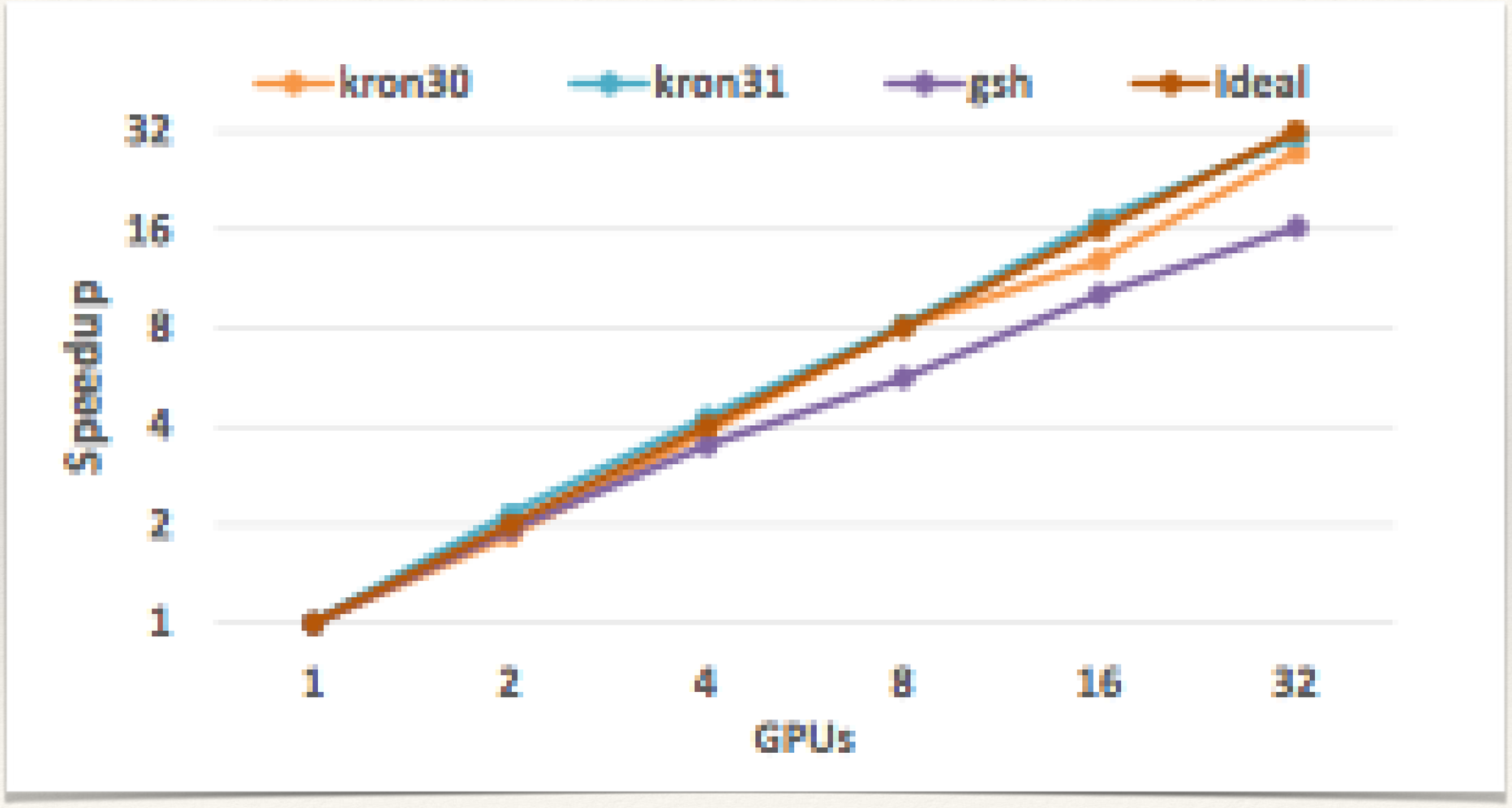
Dataset	Power (Watts)	KTEPS/Watt
Orkut	122.3	250.02
Kron-25-16	130.8	55.6
Twitter MPI	118.7	94.12
Friendster	127.8	140.58

TC Speedup over Serial Code

- Comparison vs. Serial Code (TC)



TC Scalability on GPU cluster



K-truss Performance

KTEMPS	Tiwtter MPI	Friendster	Orkut	Kron-25-16
k=3	5173.417266	12409.927	29730.58964	5903.168016
k=4	533.2362692	1432.29197	4523.326224	5305.076018
k=5	139.9433143	942.1993458	2137.493373	5540.857175
k=6	437.6643674	556.1231866	2319.407702	1637.307702
k=7	305.8757224	1242.695022	3865.679504	1590.729858
k=8	288.0275137	1241.843278	11199.87966	1546.016354

K-truss Performance

Dataset	Time (seconds)	MTEPS
email-EuALL	0.18	195.5
soc-Epinions1	0.946	202.5
flickerEdges	14.93	15.52
RoadNet	0.58	13.54
graph500-scale18-ef16	28.07	42.89
cit-Patents	8.45	477

GW

Acknowledgement

howie@gwu.edu
https://github.com/iHeartGraph