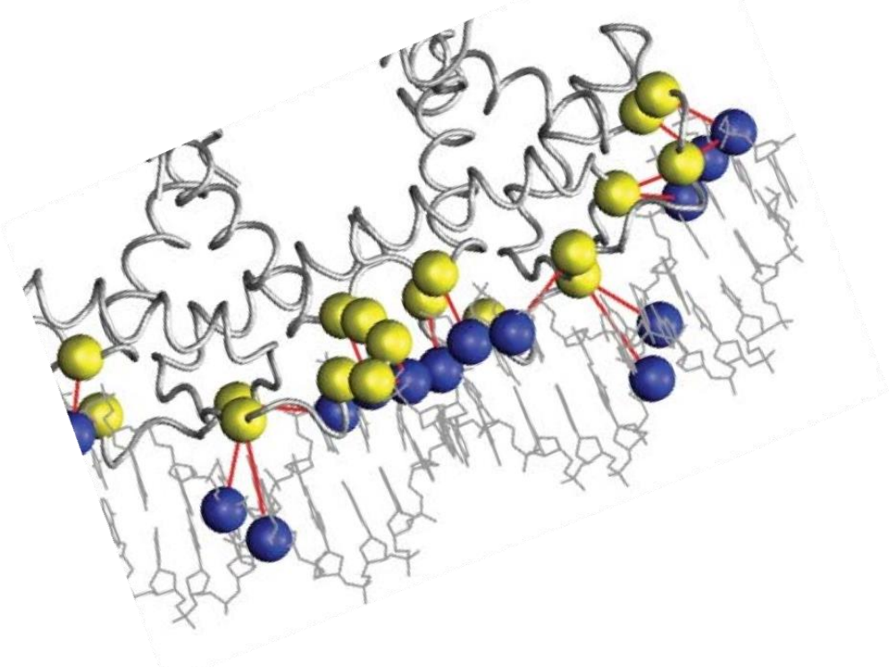
RealGraph^{GPU}: A High-Performance GPU-Based Graph Engine toward Large-Scale Real-World Network Analysis

Myung-Hwan Jang, Yunyong Ko, Dongkyu Jeong, Jeong-Min Park, and Sang-Wook Kim
Hanyang University, Seoul, Republic of Korea

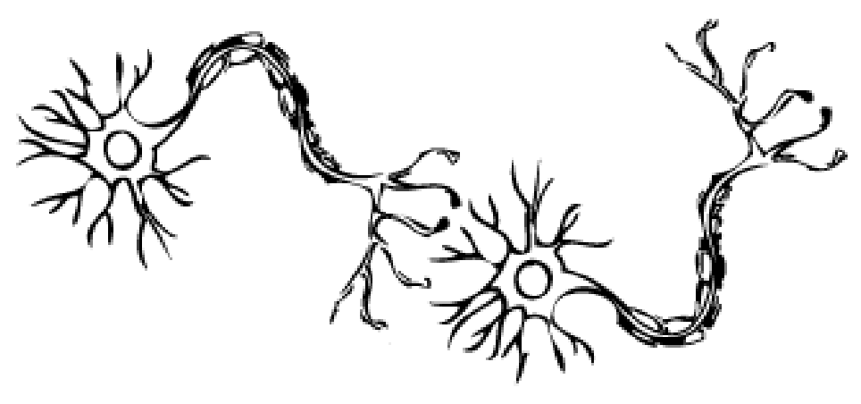
Background

- Importance of network analysis
  - ✓Many types of objects have complex relationships with each other
  - ✓*Useful information and knowledge* can be obt⁻ined by network analysis

Social networks



Protein networks

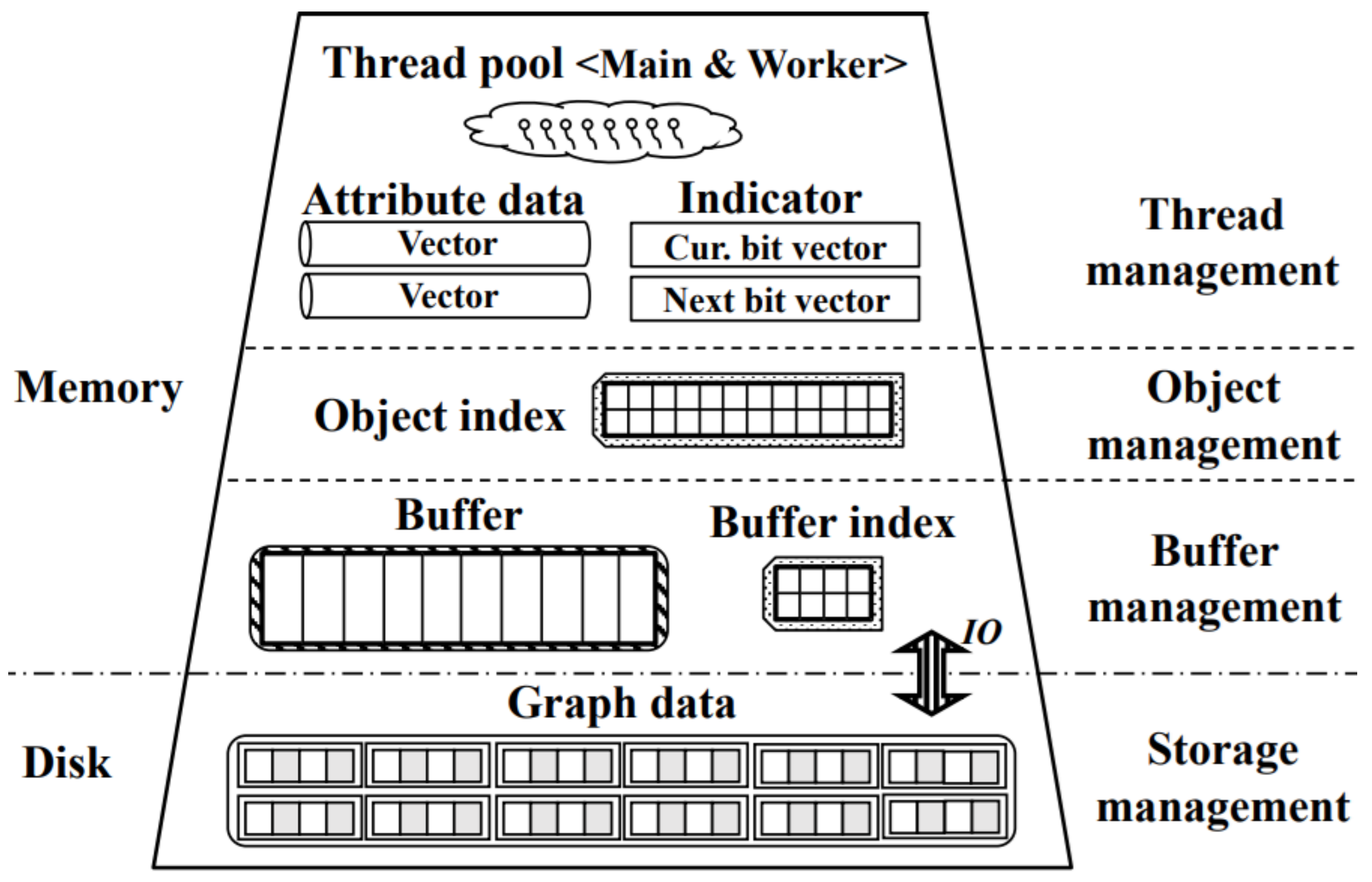


Neural networks

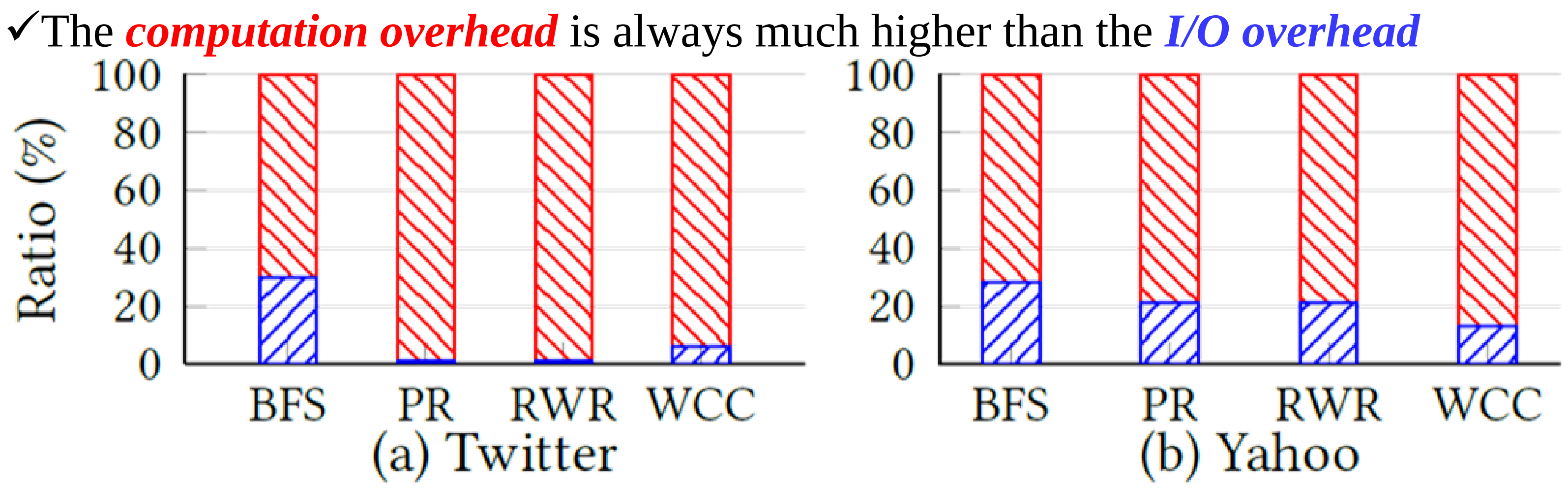
- Growing size of real-world networks
  - ✓*Efficient analysis* has become an important issue

Motivation

- RealGraph (WWW'19): *A CPU-based graph engine on a single PC*
  - ✓Processing large-scale graphs by *handling storage-to-MM I/Os efficiently*
    - ➢The main challenge occurred inevitably by a limited memory size on a single machine



- An in-depth analysis of RealGraph
  - ✓The *computation overhead* is always much higher than the *I/O overhead*



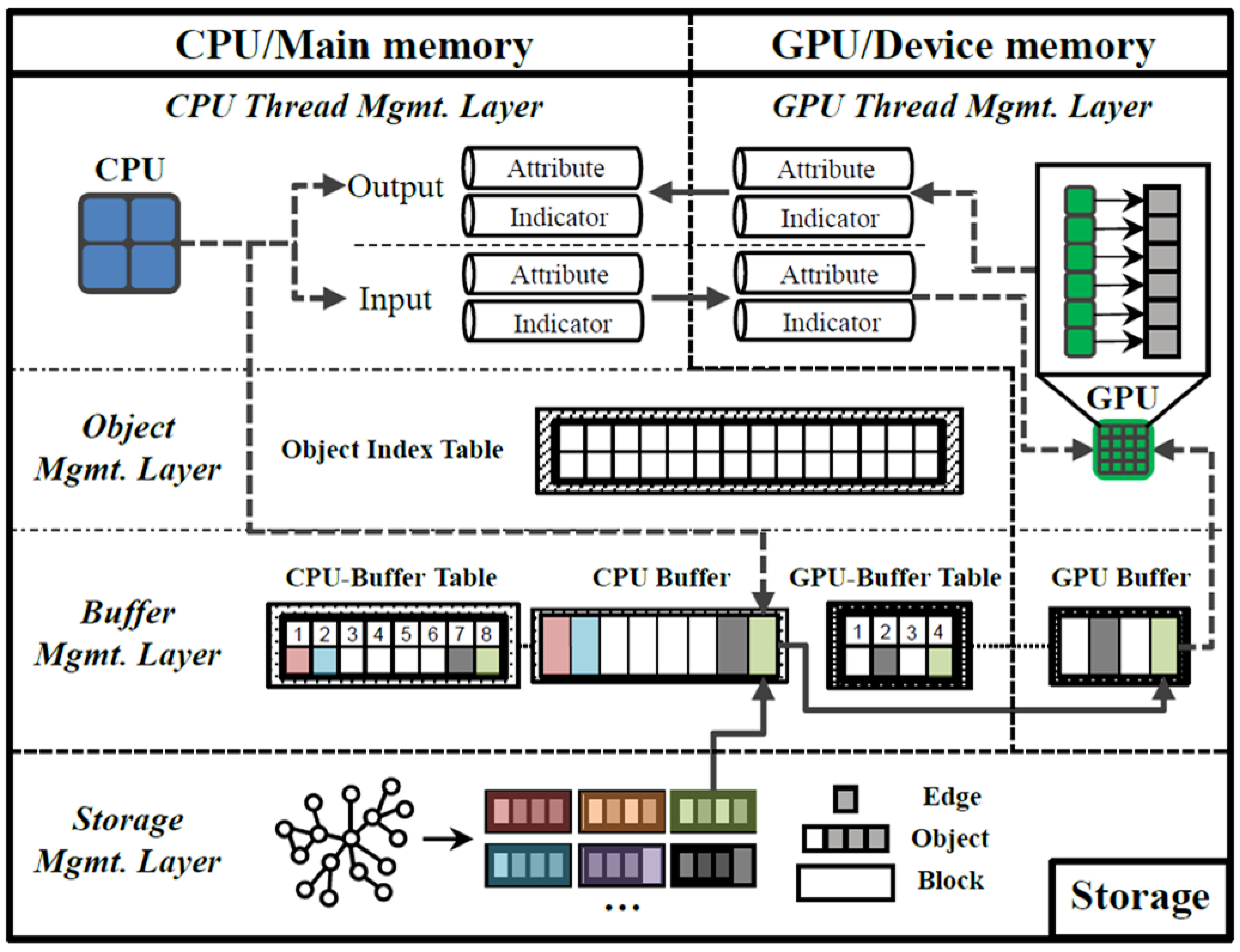
(a) Twitter (b) Yahoo

- *Utilizing a GPU for fast parallel graph processing* instead of a CPU
  - ✓*Expecting more improvement in the computation part of RealGraph*

Proposed Method: RealGraph^{GPU}

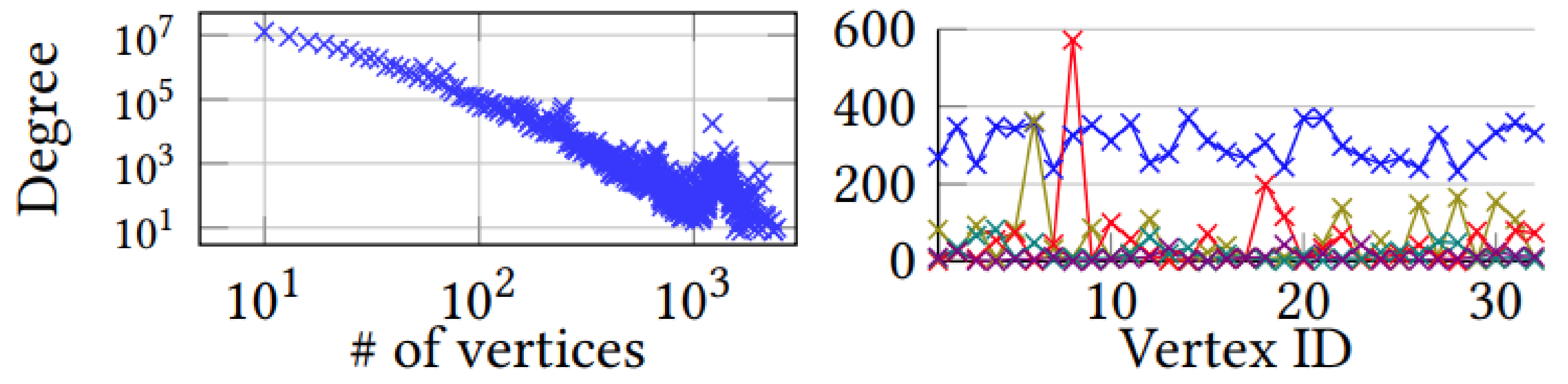
- *A GPU-based graph engine* based on original RealGraph
- A novel *5-layer architecture*

(1) Loading blocks from the storage into GPU buffer
(2) Running operations of a given graph algorithm using GPU threads



- Optimization strategies to address performance degradation

(1) **Buffer pre-checking**: *Reducing unnecessary data transfers* to GPU buffer
(2) **Edge-based workload allocation**: *Balancing workloads* among GPU threads



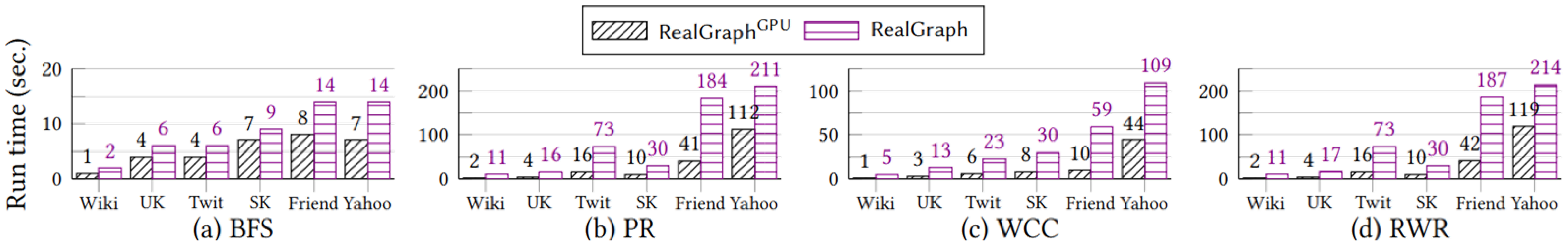
Experimental Setup

- We run our experiments on a single machine
  - ✓An Intel i7-8700K CPU with 128GB main memory and 250GB M.2 NVMe SSD
  - ✓Titan XP GPU with 12GB device memory using PCIe 3.0 interface

	Wiki	UK	Twit	SK	Friend	Yahoo
Nodes	12M	39M	61M	50M	68M	1.4B
Edges	370M	930M	1.4B	1.9B	2.5B	6.6B
Size	5.7GB	16GB	24GB	32GB	44GB	114GB

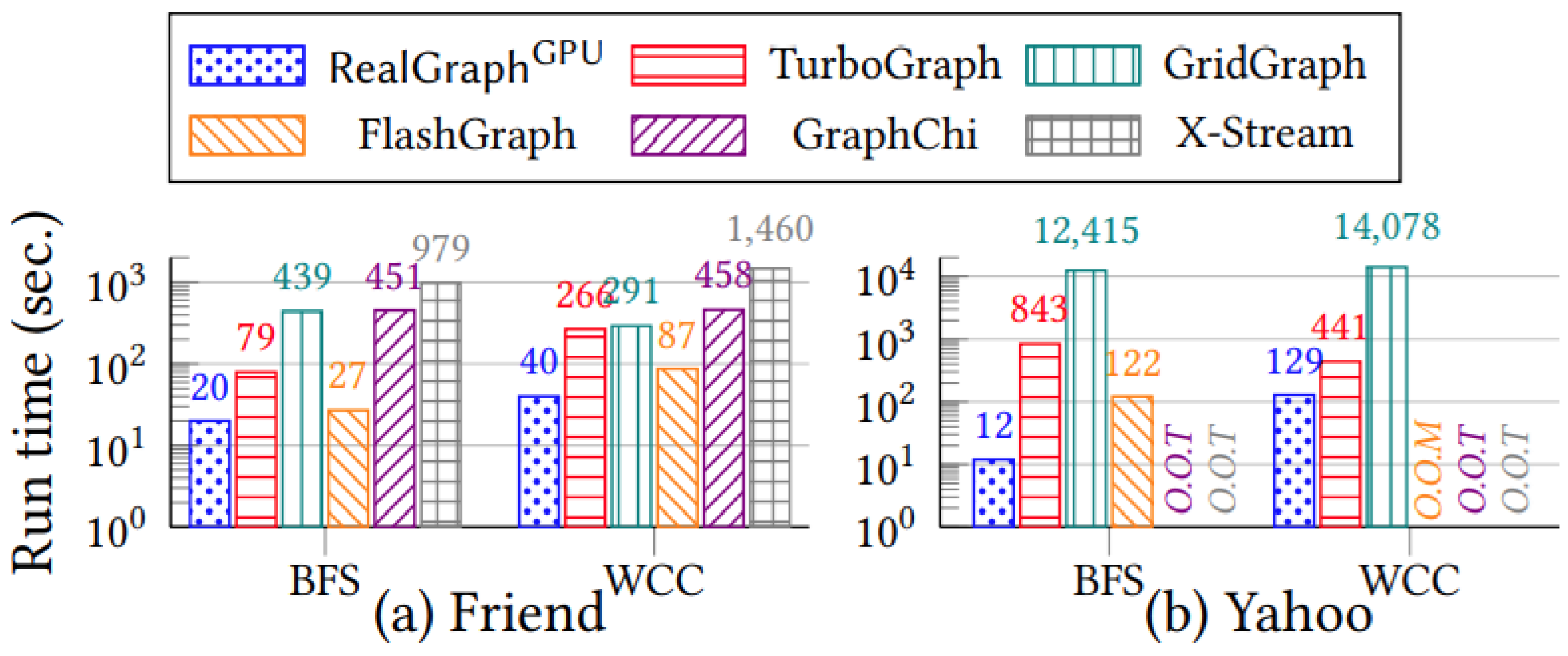
Evaluation

EQ (1): How much does RealGraph^{GPU} improve RealGraph in terms of the performance of graph processing?



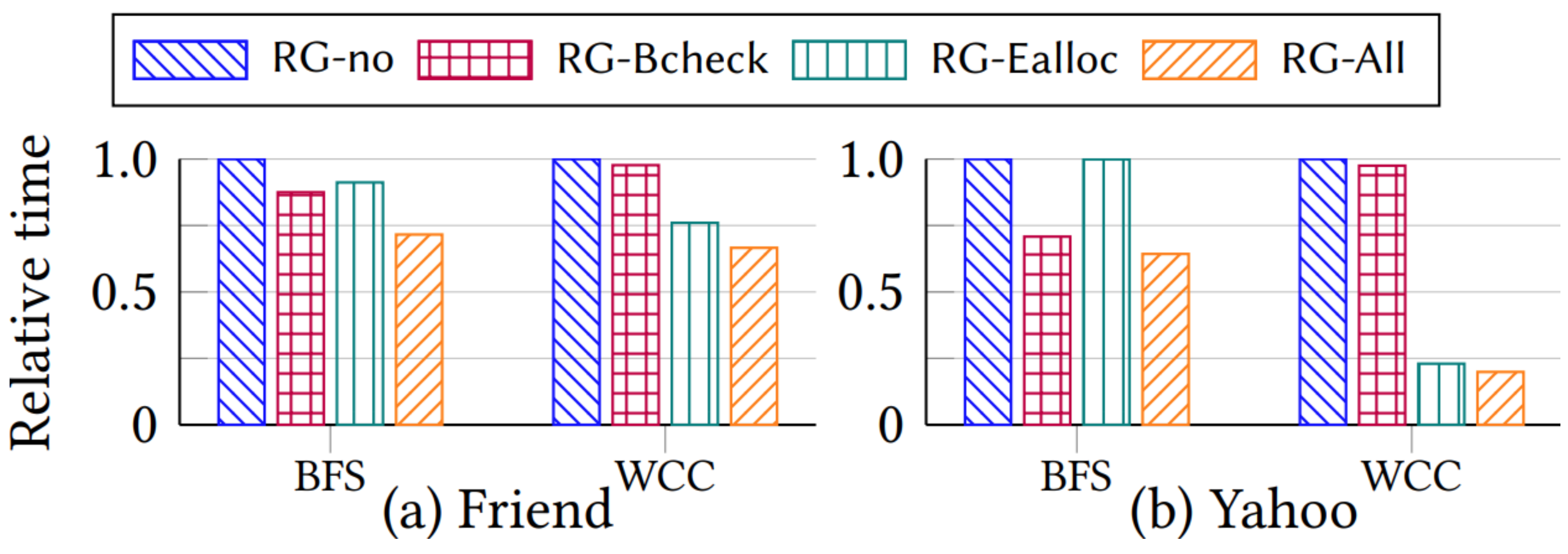
(a) BFS (b) PR (c) WCC (d) RWR

EQ (2): Does RealGraph^{GPU} provide the performance better than the existing state-of-the-art graph engines?



(a) Friend (b) Yahoo

EQ (3): Are the optimization strategies effective in improving the performance of RealGraph^{GPU}?



(a) Friend (b) Yahoo