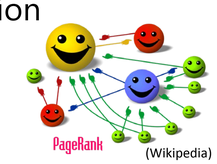
Characterizing Data Locality in Parallel Graph Algorithms

Nicole Rodia and Kunle Olukotun

Stanford University

Graph Analytics

- Important component of data mining, machine learning, and scientific computation
- Graph algorithms are:
  - Memory intensive
  - Expensive, e.g. O$(n + m)$
- **Poor performance on multicore CPUs**



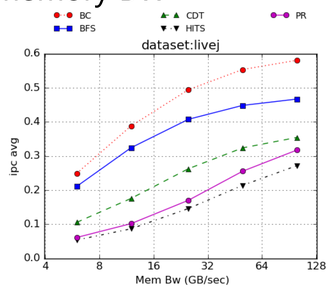
(Wikipedia)

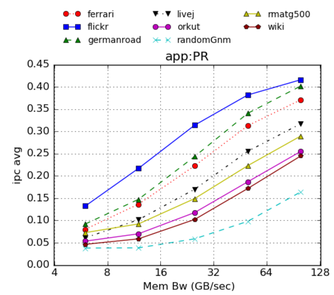
Simulation-based Analysis

- Algorithms implemented in Green-Marl DSL [1]
- Social network, web link, road, FE mesh, and synthetic graphs datasets
- Multiprocessor simulation with multi-level memory hierarchy (zsim) [2]

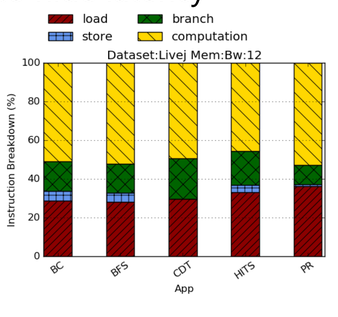
The Memory Wall

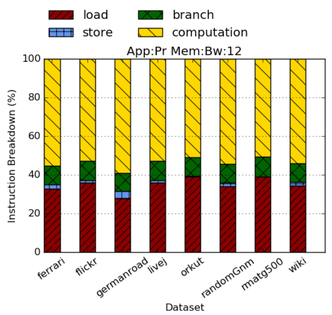
- Increasing gap between fast computation and slow data access
- Modern machines use large multi-level caches to compensate for limited memory bandwidth
- **Execution time is dominated by data access**
- Random access and large data size heavily utilize memory BW





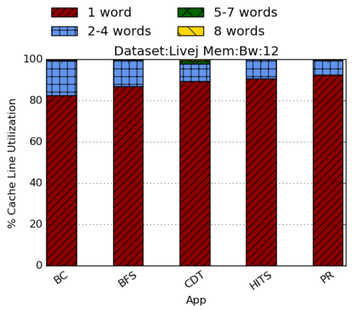
- Low computation-to-memory access ratio unable to hide latency

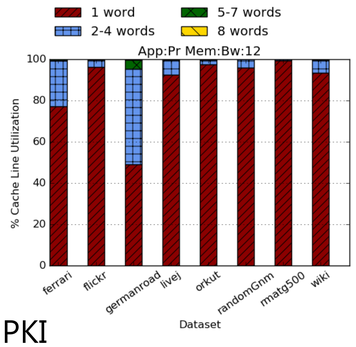




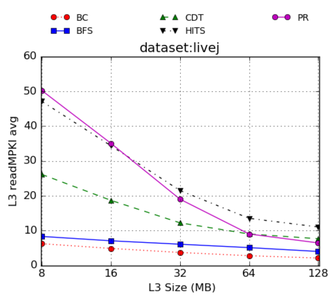
Data Cache Performance

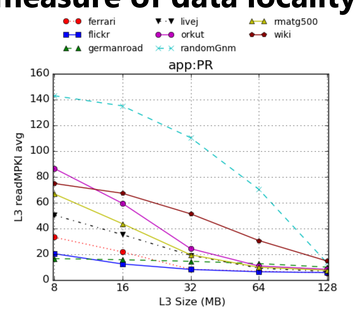
- Cache lines poorly utilized → wastes memory BW





- ↑ cache size → ↓ cache MPKI
- Dataset size and structure affect cache MPKI
- **Miss rates not reliable measure of data locality**

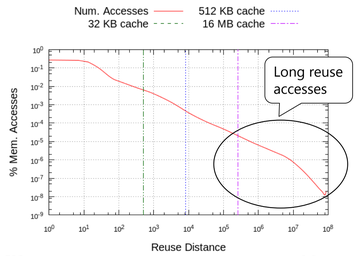


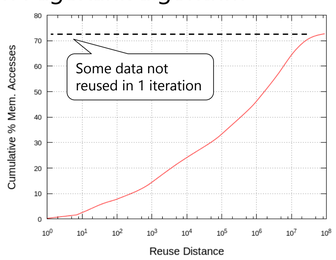


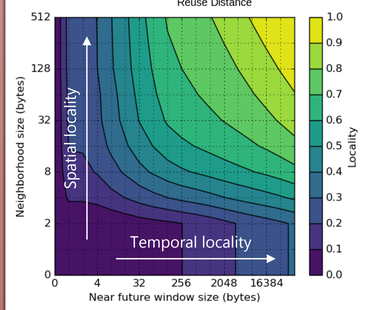
Data Locality

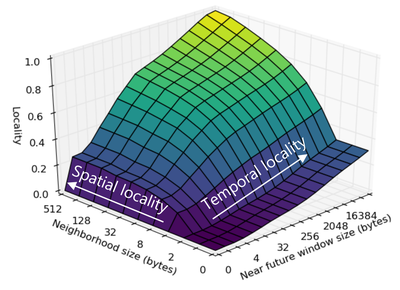
- Temporal locality: access same location again in future
- Spatial locality: access nearby location in future
- Long reuse accesses poor fit for LRU replacement policy

LiveJournal dataset on PageRank algorithm



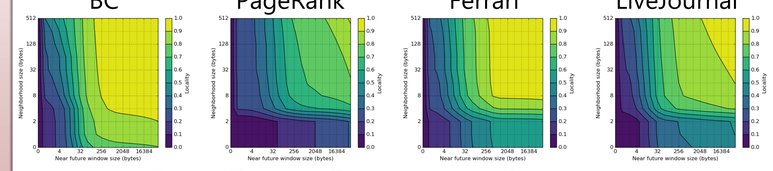






- **Algorithm and dataset properties dictate data locality**

Average locality score [3]

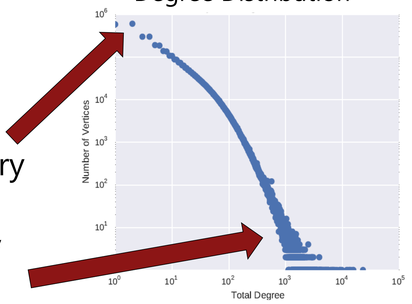


Per-Data Structure Locality

Informatics Graph Properties

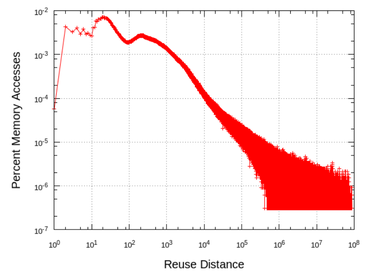
- Small-world: O(log n) diameter
- Scale-free: Power-law degree distribution
- Many vertices with very small degree
- Few vertices with very large degree

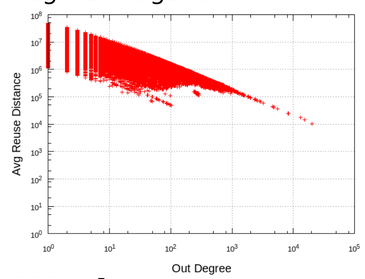
LiveJournal social network Degree Distribution



- **Average reuse distance correlates with vertex degree** for data structures with non-uniform access patterns

PageRank values data structure
LiveJournal dataset on PageRank algorithm





Future Work

- Leverage algorithm, data structure, and graph dataset properties for a more effective cache replacement policy that better captures data locality and thus improves performance
- Incorporate fine-grained data access to reduce unused cache space and improve spatial locality

References

[1] S. Hong, H. Chafi, E. Sedlar, and K. Olukotun, "Green-Marl: A DSL for easy and efficient graph analysis," *ASPLOS* '12.

[2] D. Sanchez and C. Kozyrakis, "Zsim: Fast and accurate microarchitectural simulation of thousand-core systems," *ISCA* '13.

[3] S. Gupta, P. Xiang, Y. Yang, and H. Zhou, "Locality principle revisited: A probability-based quantitative approach," *IPDPS* '12.