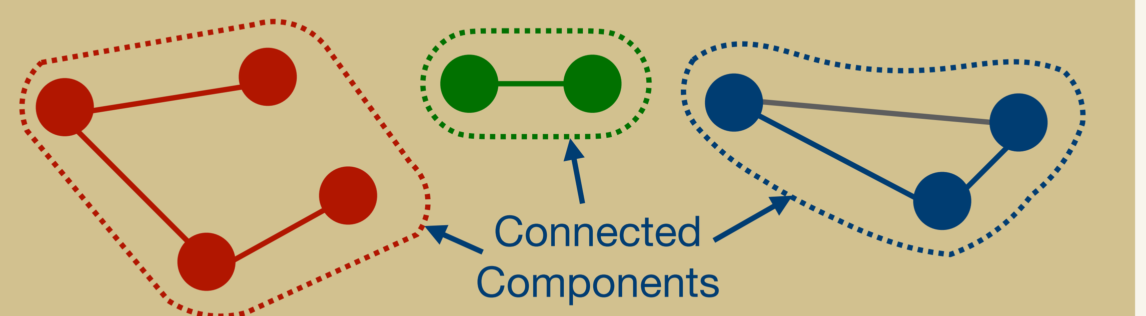
When Label Propagation outperforms BFS in Breadth-First Graph Traversal at scale

Kalsuda Lapborisuth, *Srinivas Aluru (Advisor)*

Breath-First Graph Traversal

Computes a BFS ordering for the **entire graph** across all connected components (CC).



Applications:

Triangle Counting	Graph Compression
Computing Shortest Path	k-clique Counting
Connected Components	Graph Reordering

Issue: Traversing with BFS is **work optimal** but **lacks scalability** in graphs with a large number of connected components.

Label Propagation (LP)

Initialize: Place each vertex in its own partition

Each iteration:

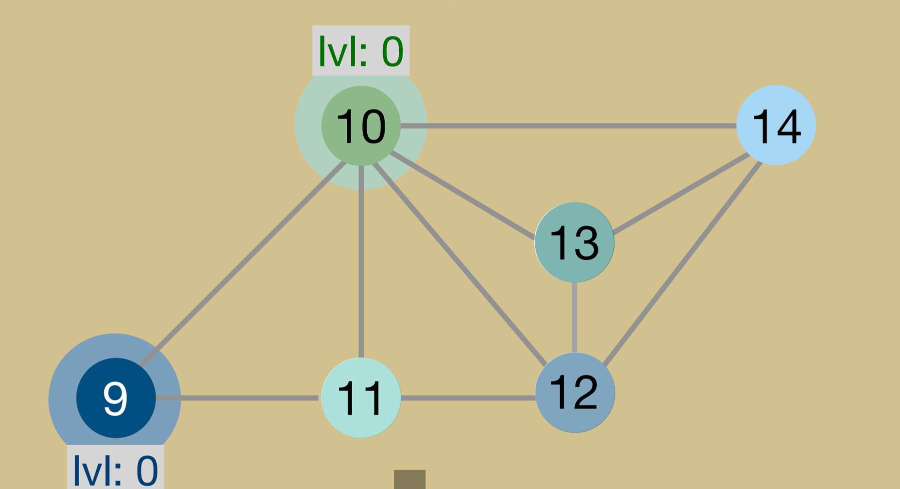
(1) Exchange partition ID & level with other PE

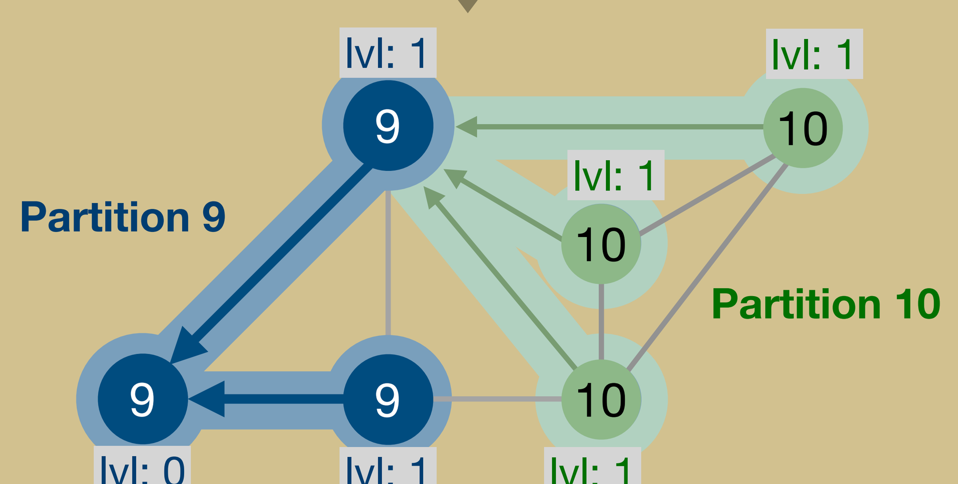
Synchronization (MPI_AlltoAllv)

(2) Update partition ID, level, and parents via:

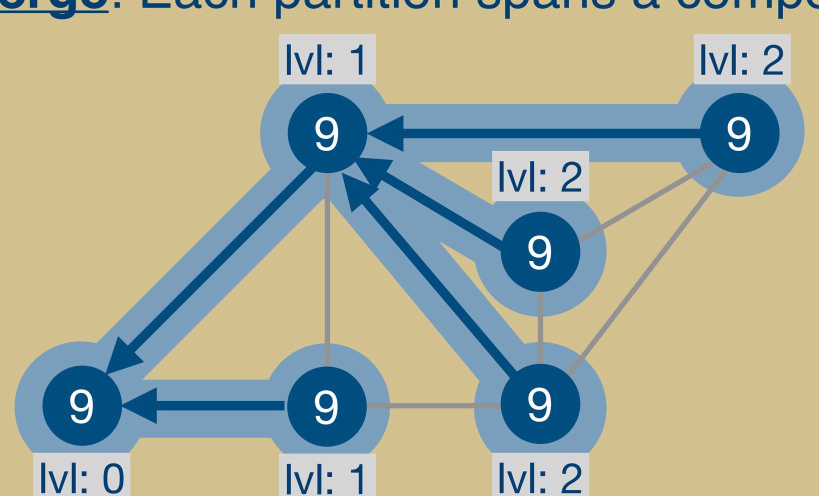
$$ID_i[v] = \min_{u \in N(v)} ID_{i+1}[u]$$

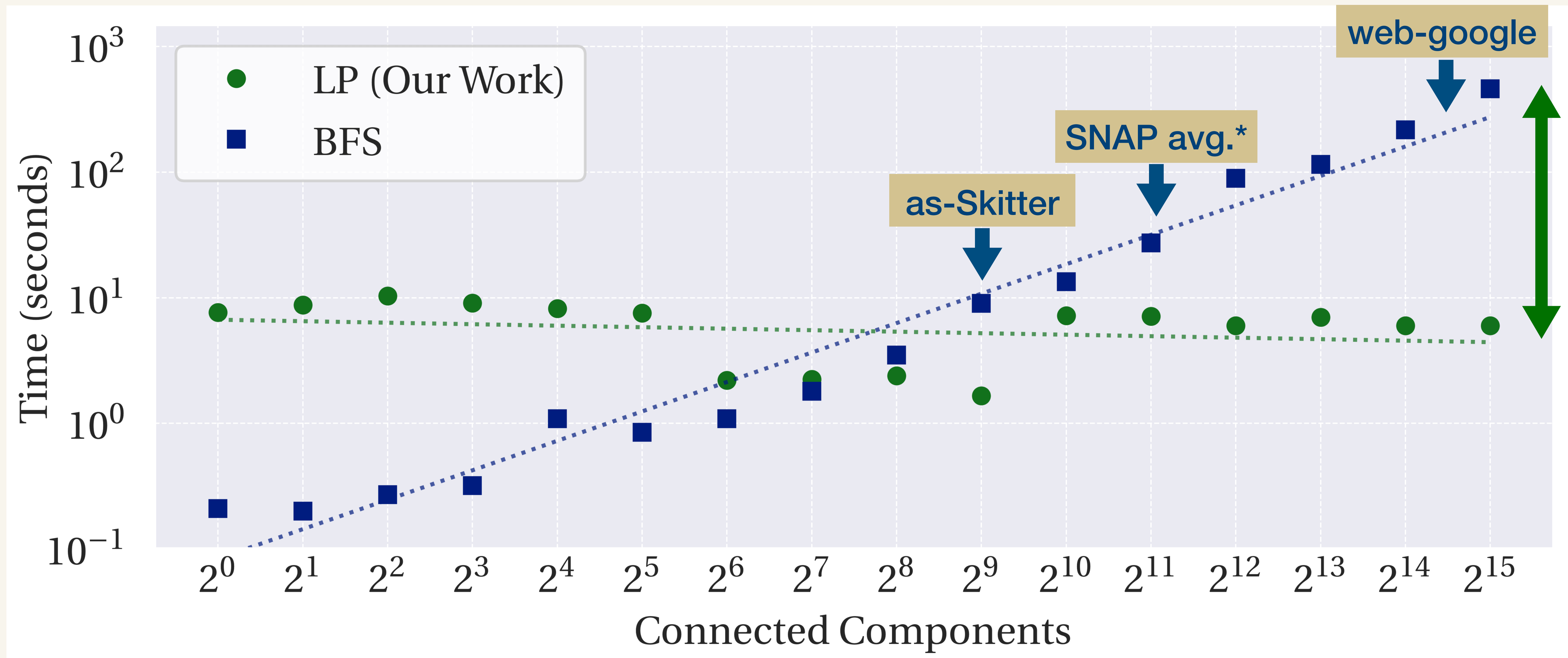
$$Lvl[v] = Lvl_{i+1}[u] + 1$$





Converge: Each partition spans a component



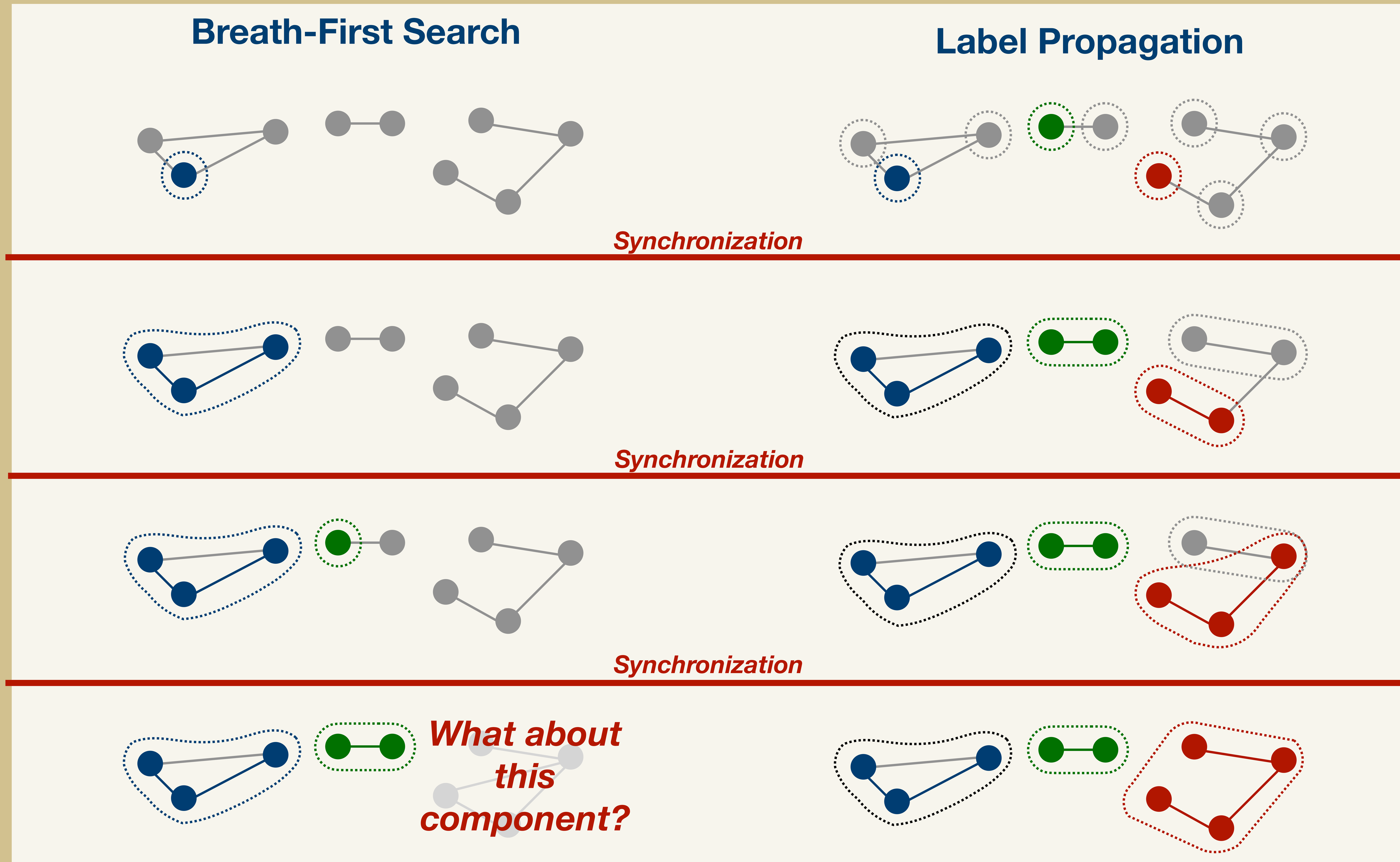


77x Speedup

using ***label propagation*** to traverse all connected components at once.

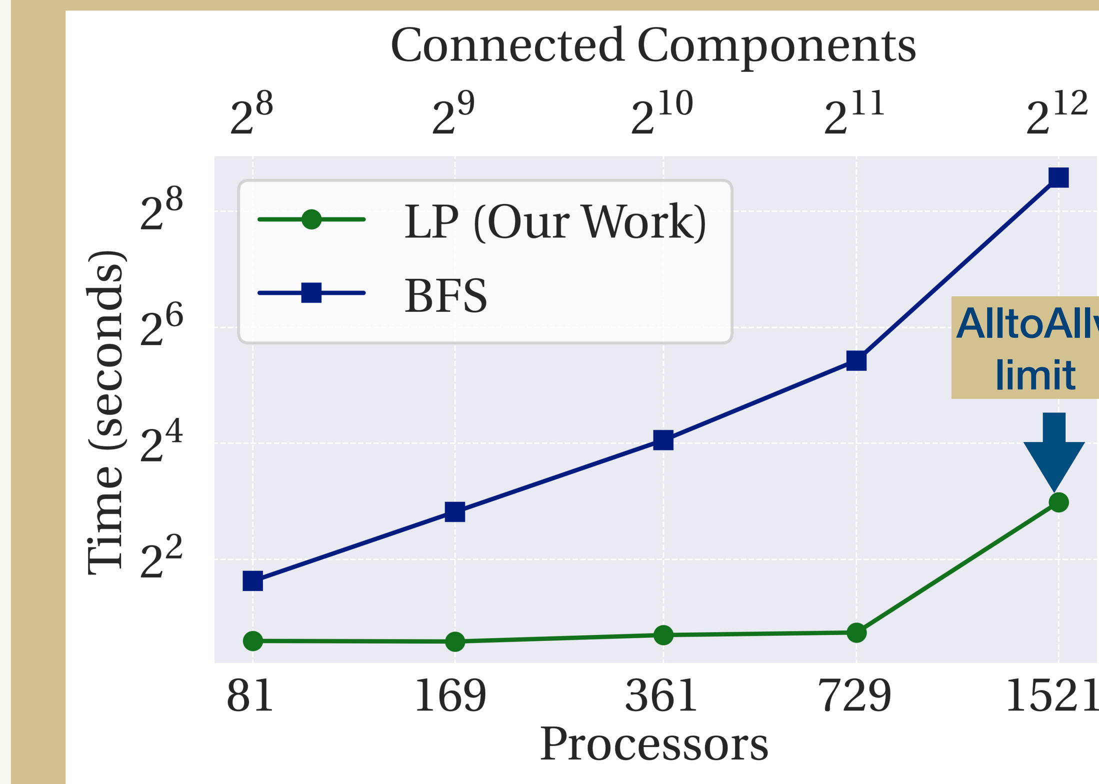
30,000x less synchronizations,

despite more work and communication.



More Work for Parallelism & Less Synchronization

Weak Scaling Experiment



Worst-Case Bound

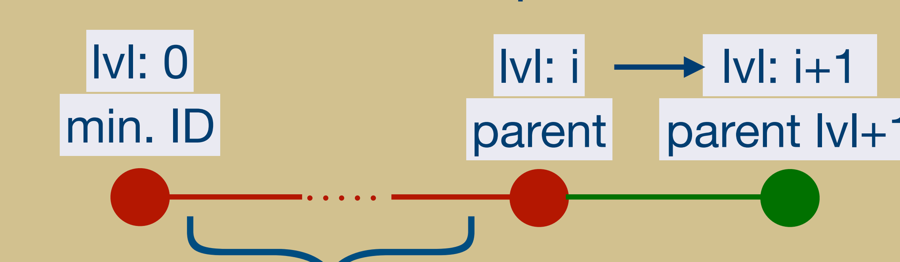
	Synchronization	Communication Volume & Work
LP	Diameter of the graph d	$\mathcal{O}(d(\|V\|+\|E\|))$
BFS	Sum of diameter of all components	$\mathcal{O}(\|V\|+\|E\|)$

Real-World Graph Examples

Graphs	|V|	|E|	d	#CC
as-Skitter	1.6M	11M	25	756
roadNet-TX	1.3M	1.9M	1054	13890
Human Metagenome	2.0B	2.0B	3989	1.0M
Twitter	52.6M	2.0B	16	29533
Soil (Peru)	53.7B	523.6M	2463	7.6M

LP Trees are Breath-First

Theorem: The level of a vertex is its shortest path distance to the vertex with the minimum ID in the connected component.



shortest path of length i + edge (length 1)

Acknowledgement: *We thank Souvadra Hati and Richard Vuduc for their insightful discussions and for sharing their implementation of the Generalized-RMAT generator.*

Georgia Tech

*****SNAP Avg.:*** *The average of randomly selected 48 out of 71 SNAP graphs available in SuiteSparse Matrix Collection.*