

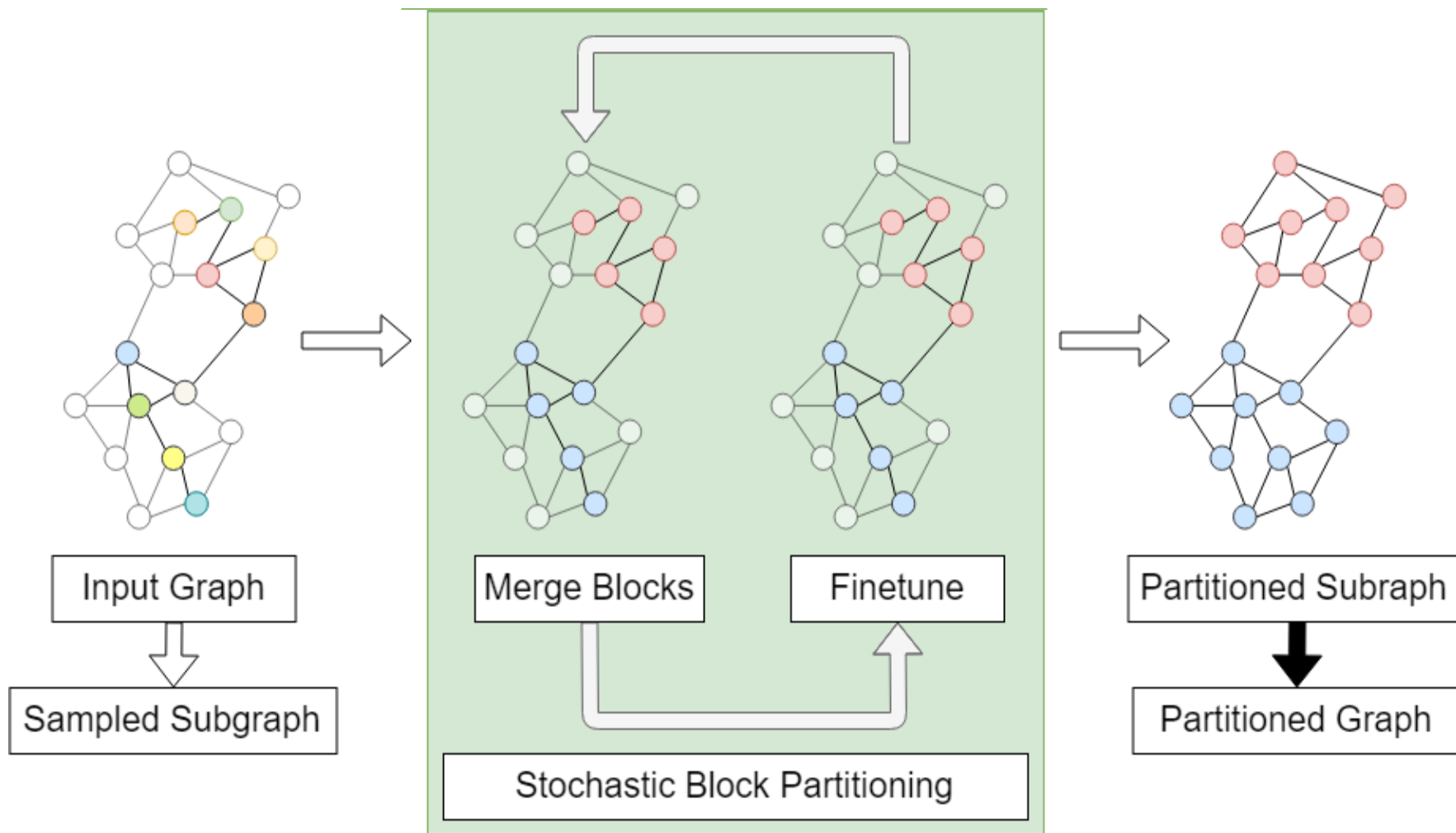
# FAST STOCHASTIC BLOCK PARTITIONING VIA SAMPLING

FRANK WANYE, VITALIY GLEYZER, WU-CHUN FENG

2019 IEEE HIGH-PERFORMANCE EXTREME COMPUTING (HPEC'19) CONFERENCE

SEPTEMBER 25, 2019

# INTRODUCTION: STOCHASTIC BLOCK PARTITIONING



We speed up stochastic block partitioning via sampling without compromising community detection results

# EXPERIMENT OVERVIEW

## ALGORITHMS

- Uniform Random (UR)
- Random Node Neighbor (RNN)
- Degree Weighted (DW)
- Random Walk (RW)
- Random Jump (RJ)
- Forest Fire (FF)

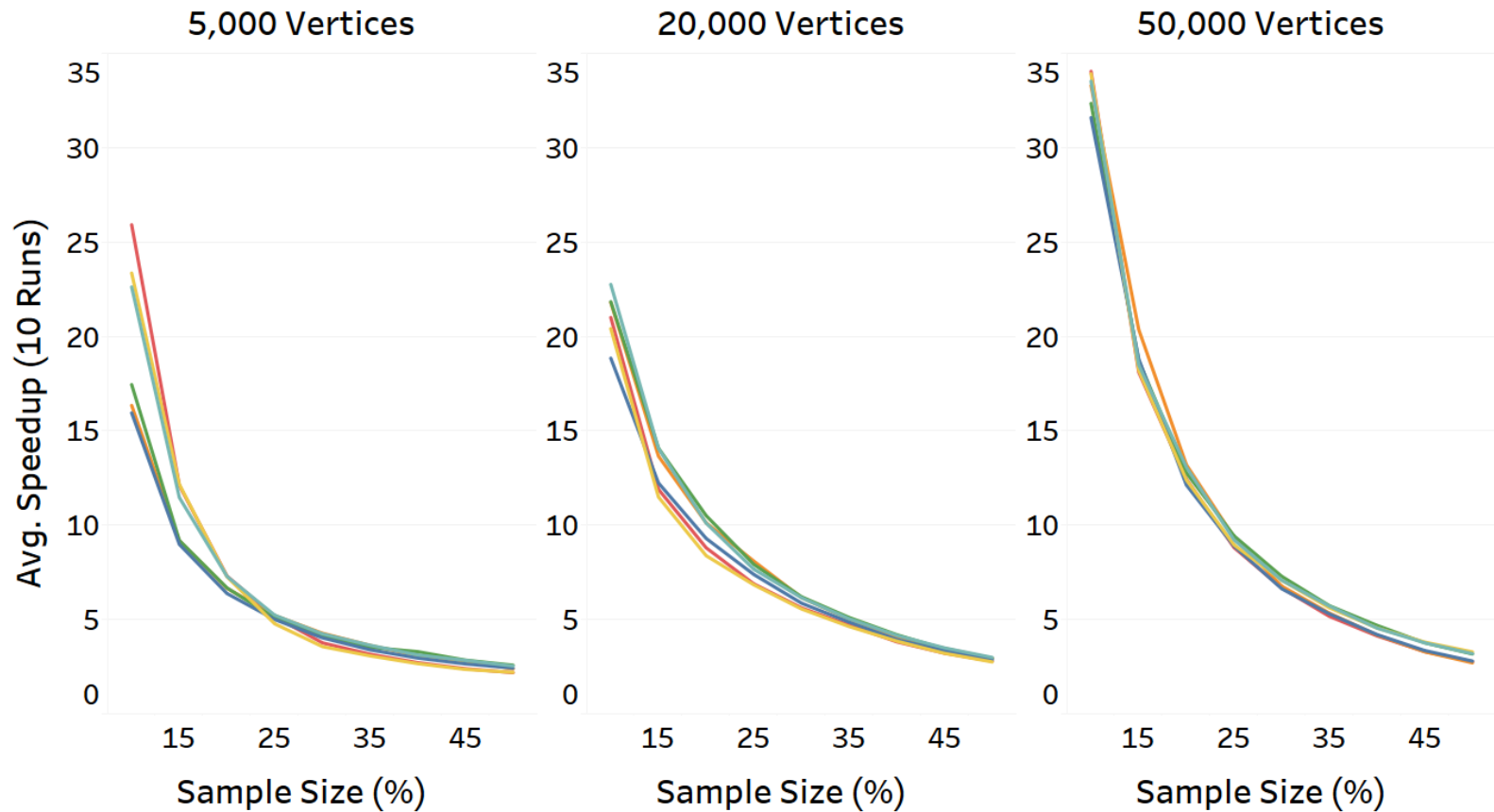
## VARIABLES

- Graph size (5K, 20K & 50K vertices)
- Difficulty of partitioning
- Sample size (10% - 50%)

## METRICS

- $F1\ Score = \frac{2 * precision * recall}{precision + recall}$
- Speedup
- Others (refer to paper)

# RESULTS: SPEEDUP ON THE BENCHMARK (NO SAMPLING)



## Sampling Algorithm

- Degree Weighted (DW)
- Random Jump (RJ)
- Random Walk (RW)
- Forest Fire (FF)
- Random Node Neighbor..
- Uniform Random (UR)

At 50% sample size, all algorithms performed similarly. At 10% sample size, generally performed best