

# Cut to Fit: Tailoring the Partitioning to the Computation

Iacovos G. Kolokasis, Polyvios Pratikakis

Foundation for Research and Technology Hellas (FORTH)  
Computer Science Department, University of Crete

## Partitioning and Placement

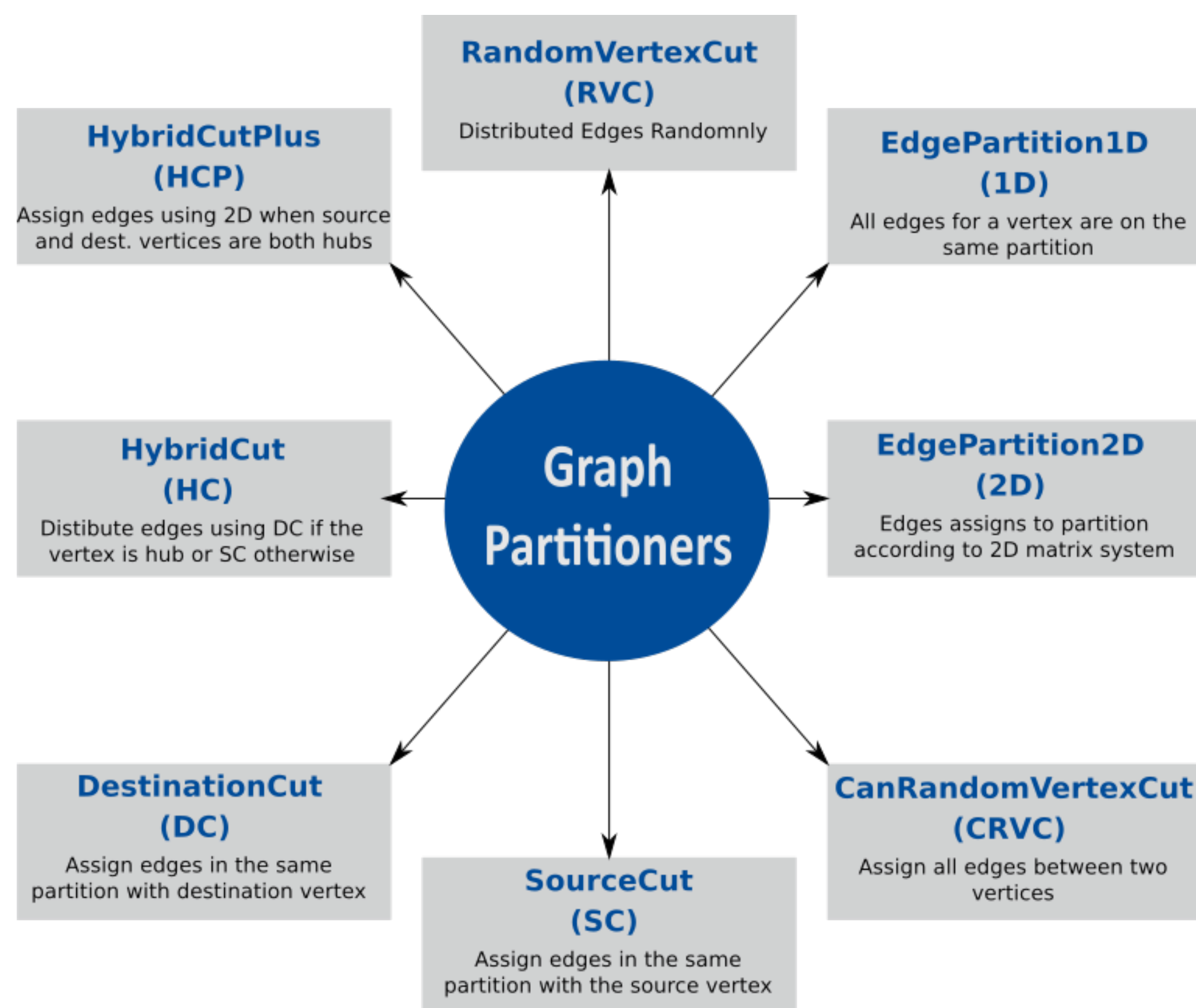
- There is no single optimal partitioner for all problems
- There is no single partitioning metric which is always correlated with computation performance
- We propose PARSEL, a heuristic way to select a partition strategy

## Dataset Analysis

Dataset	Vertices	Edges	Type	Size on Disk
web-wikipedia-link-fr	4.9M	113.1M	Power-Law	1.6G
soc-twitter-2010	21.2M	265.0M	Power-Law	4.4G
road-road-usa	23.9M	28.8M	Low-Degree	469.7M
soc-sinaweibo	58.6M	261.3M	Long-Tailed	3.8G
socfb-uci-uni	58.7M	92.2M	Long-Tailed	1.5G

Characteristics of datasets.

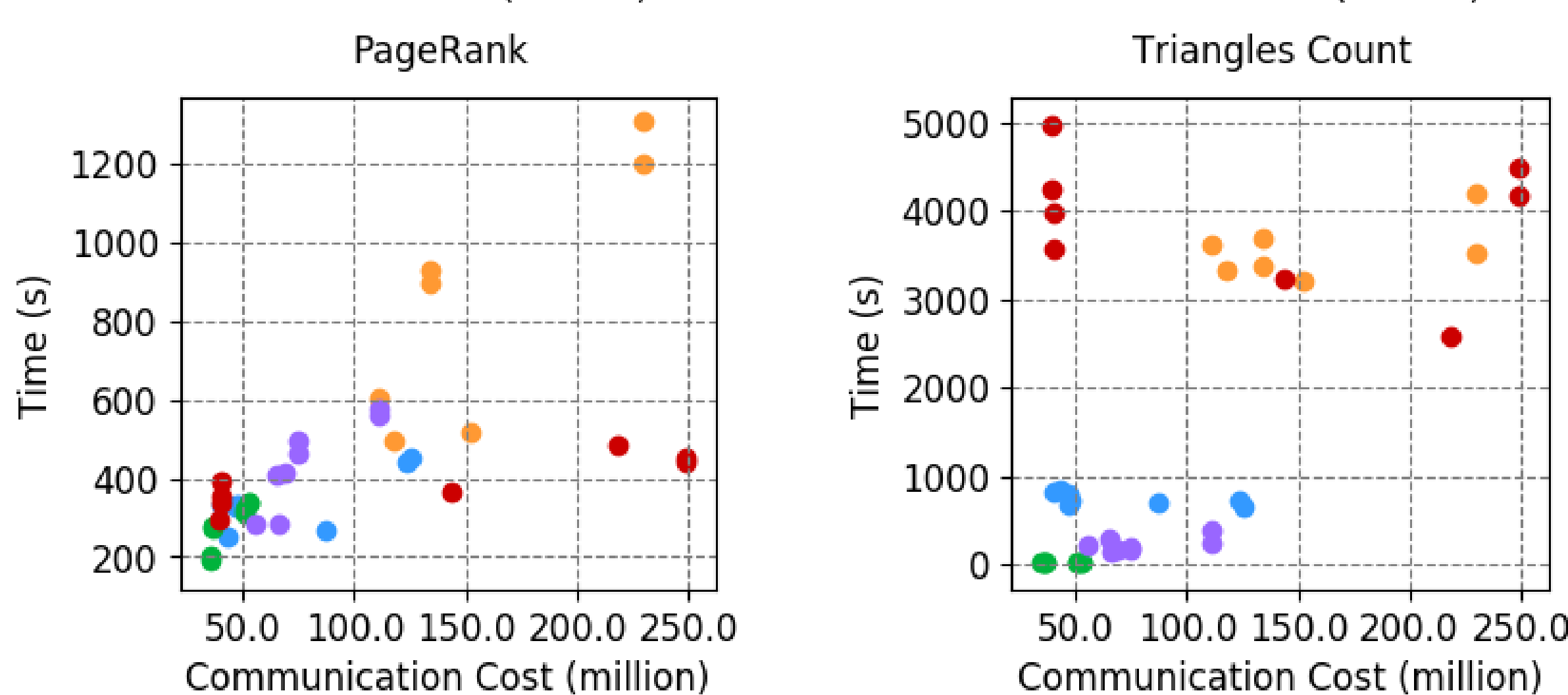
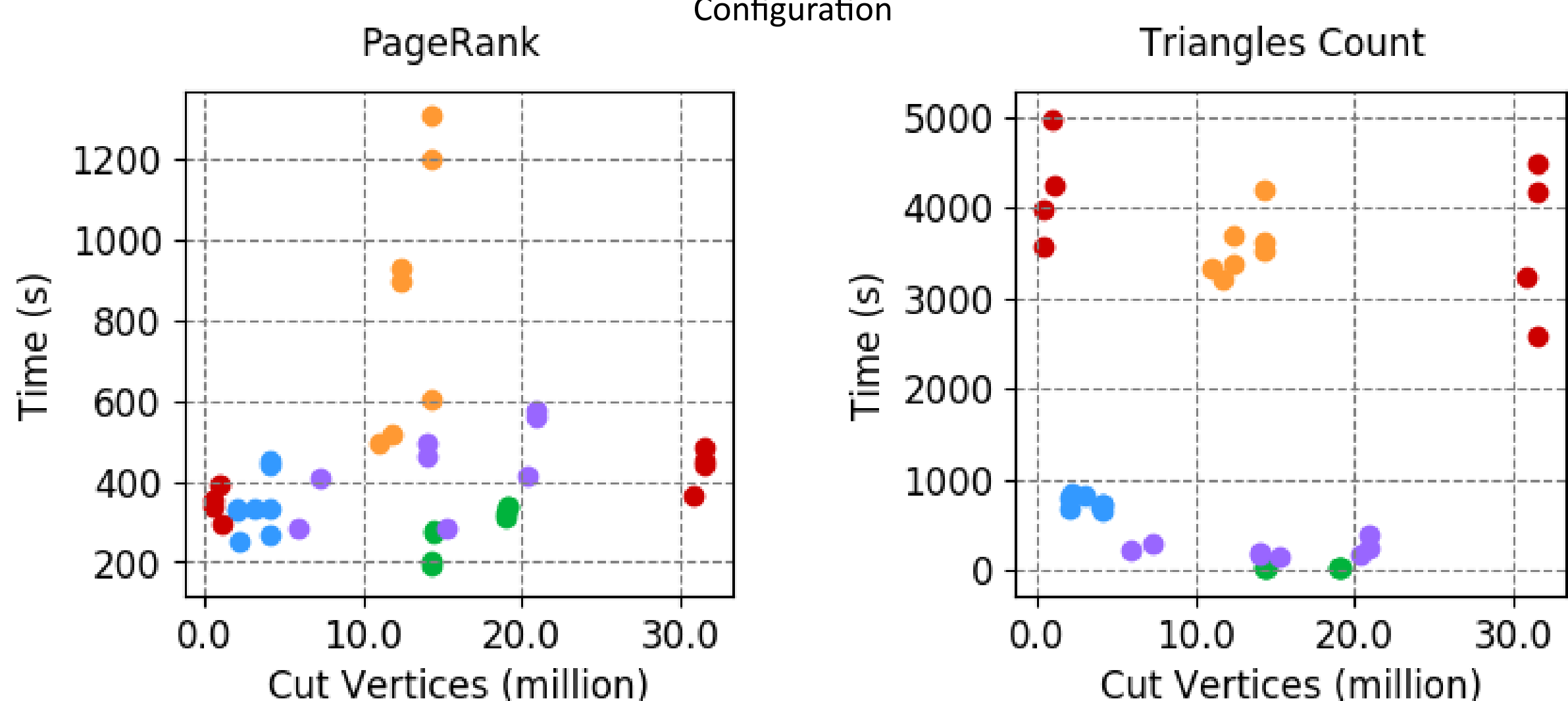
## Graph Partitioners



## Partition Metrics As Performance Predictor

Cluster	Instance	Cores	Memory	CPUs
Master	1	32	256GB	Intel(R) Xeon(R) E5-2630 CPUs
Workers	4	32	256GB	Intel(R) Xeon(R) E5-2630 CPUs

Configuration

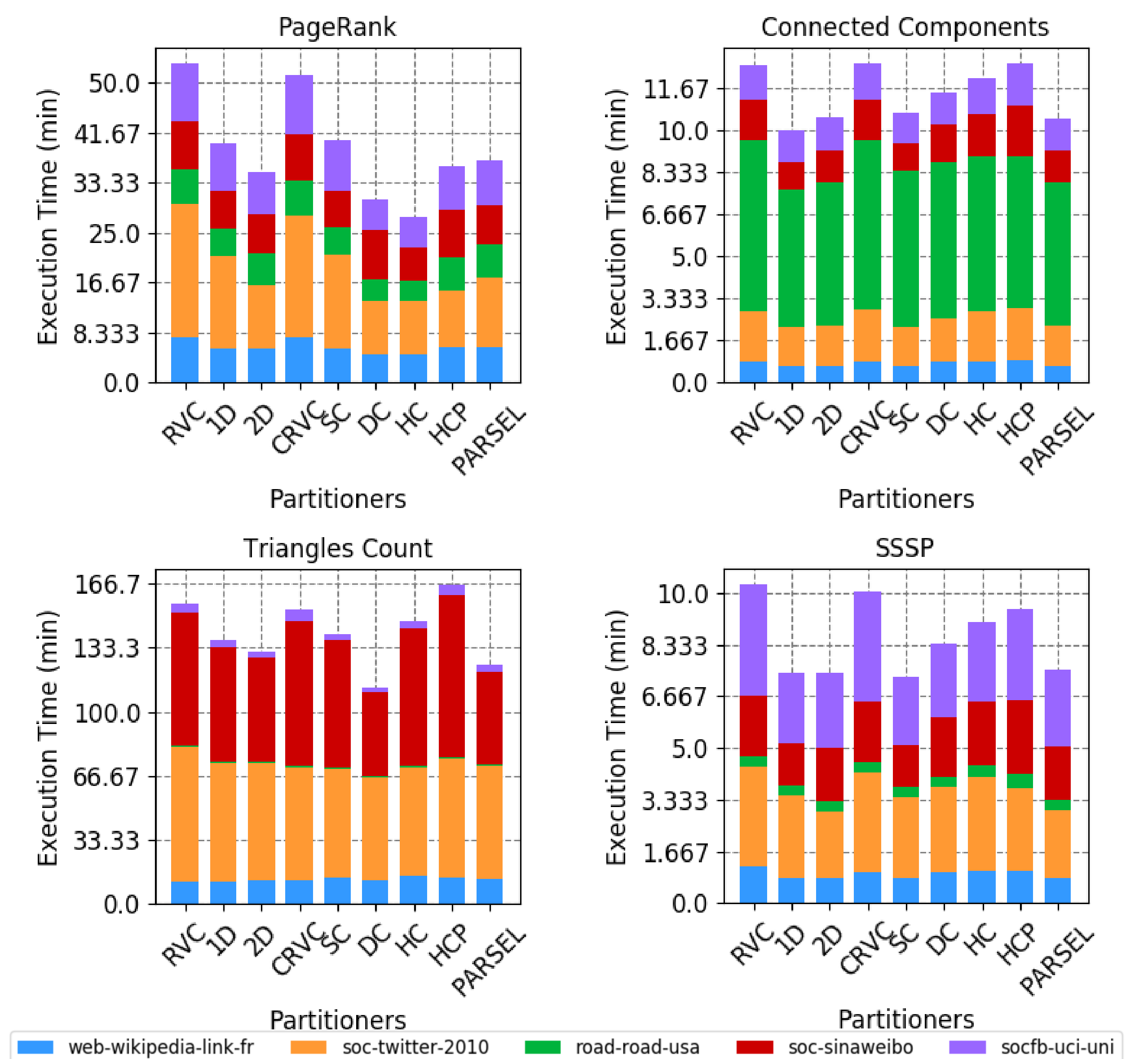
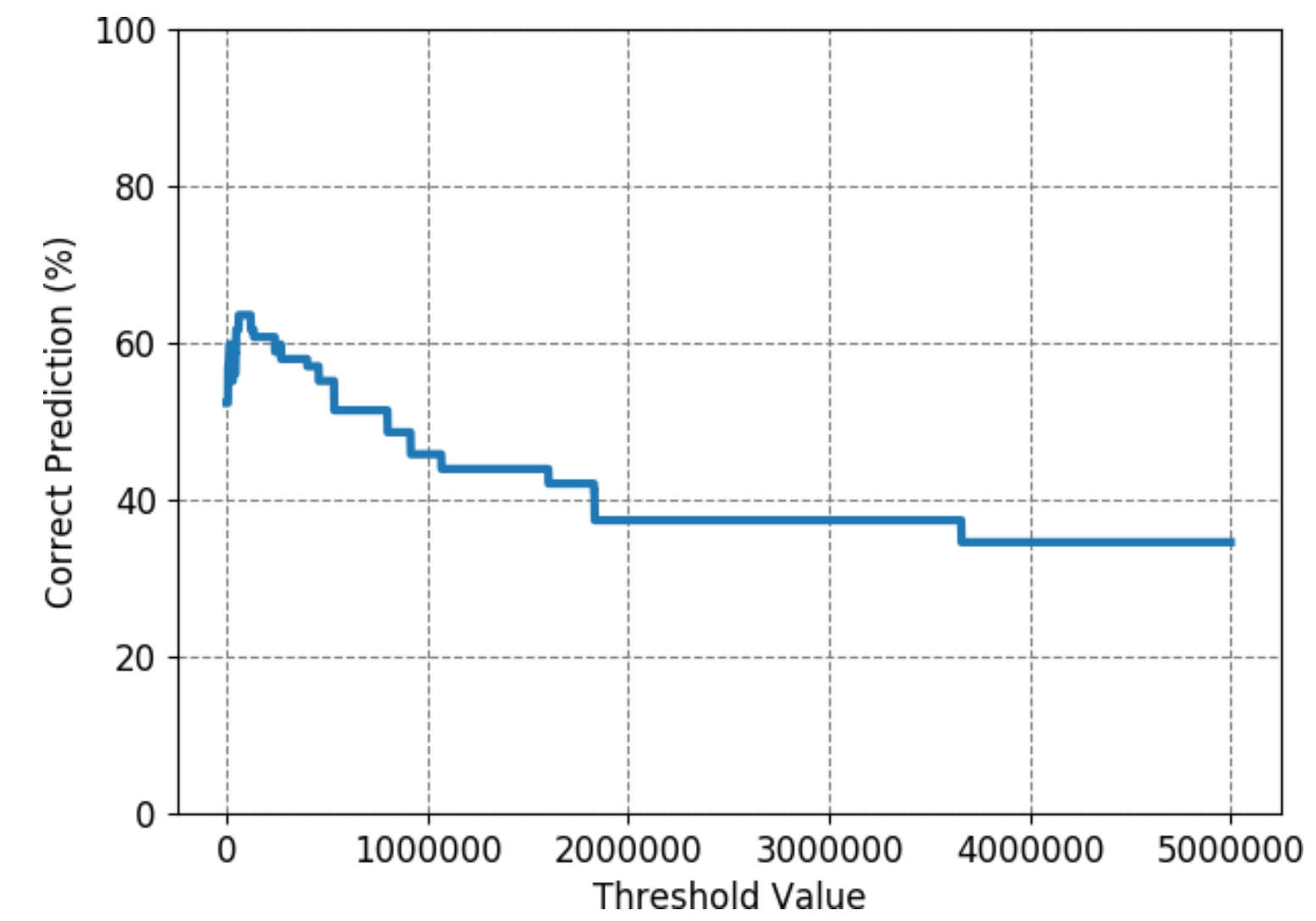


• web-wikipedia-link-fr • road-road-usa • soc-sinaweibo • socfb-uci-uni  
• soc-twitter-2010

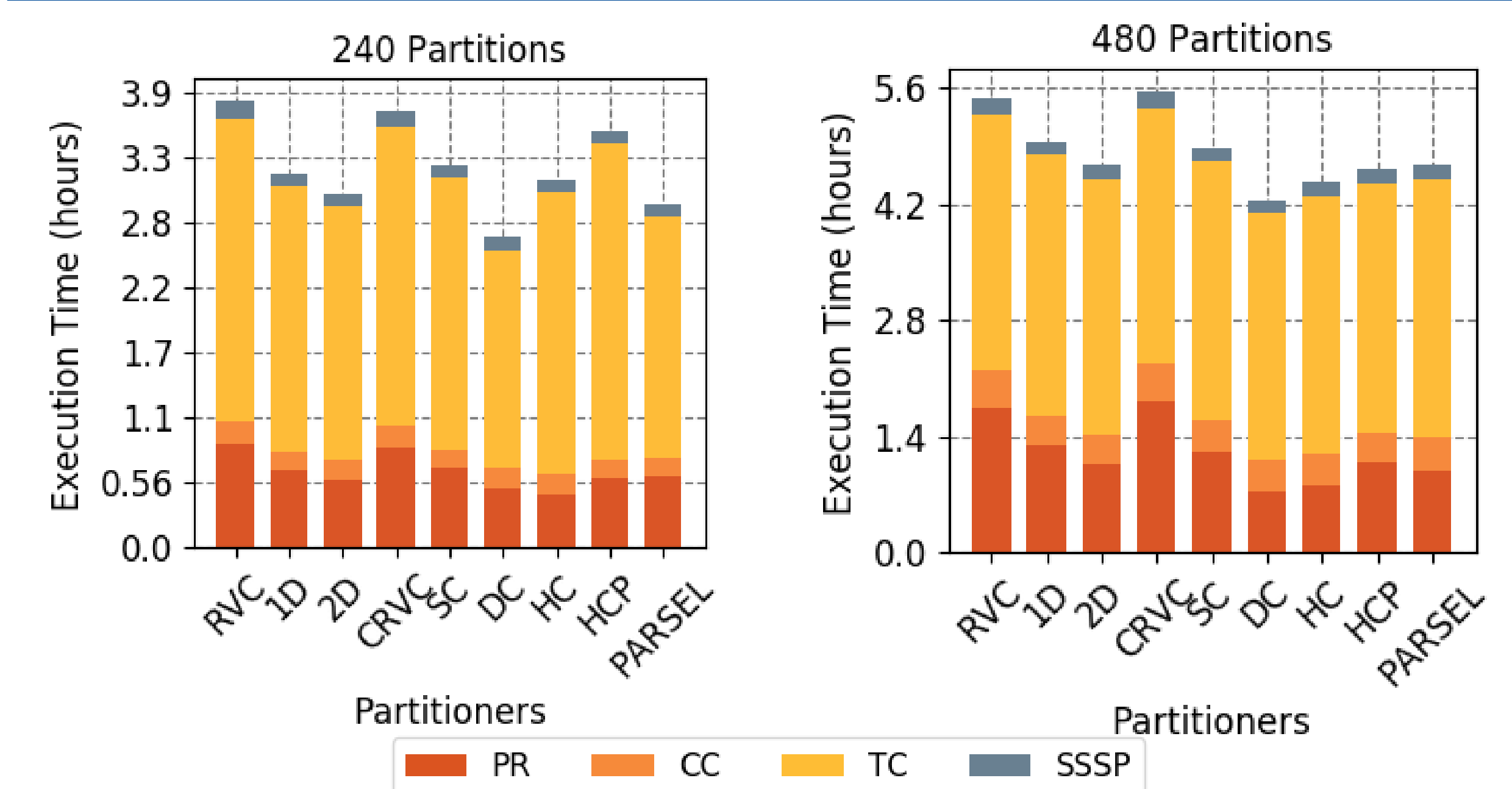
- Cut Vertices and Communication Cost are better predictors of execution time for PageRank but not for Triangles Count

## PARSEL

- PARSEL chooses between 2D and DC, on average the most efficient partition strategies
- We trained the threshold value for PARSEL
- PARSEL's decision metric is based on the ratio of total number of edges and total number of partitions



## Partitioners Over Analytics Workflow



## Conclusions

- Distributed graph analytics frameworks efficiency is highly dependent on the partitioning strategies
- No single metric is a good predictor of workload execution time
- Dynamic partitioner selection can better tolerate different computations, datasets and resource configuration compared to complex partitioners
- Re-partitioning the graph using a fast dynamically selected partitioner at each step in analytic workflow may outperform an optimized partitioner
- There is a trade-off between ingress time and computation time for the selection of a partitioner