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

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Vertices</th>
<th>Edges</th>
<th>Type</th>
<th>Size on Disk</th>
</tr>
</thead>
<tbody>
<tr>
<td>web-wikipedia-link-fr</td>
<td>4.9M</td>
<td>113.1M</td>
<td>Power-Law</td>
<td>1.6G</td>
</tr>
<tr>
<td>soc-twitter-2010</td>
<td>21.2M</td>
<td>205.0M</td>
<td>Power-Law</td>
<td>4.4G</td>
</tr>
<tr>
<td>road-road-usa</td>
<td>23.9M</td>
<td>28.8M</td>
<td>Low-Degree</td>
<td>469.7M</td>
</tr>
<tr>
<td>soc-sinaibo</td>
<td>58.5M</td>
<td>261.3M</td>
<td>Long-Tailed</td>
<td>3.8G</td>
</tr>
<tr>
<td>socfb-uci uni</td>
<td>58.7M</td>
<td>92.2M</td>
<td>Long-Tailed</td>
<td>1.5G</td>
</tr>
</tbody>
</table>

Characteristics of datasets.

Graph Partitioners

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.

Graph Partitioners As Performance Predictor

Cluster Metrics As Performance Predictor

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.

kolokasis@ics.forth.gr