Multilayer networks in GraphLab

An open source project study
GraphLab

Execution engine: PowerGraph
Used for graph processing
  - vertex-centric functions
  - distributed data graph
  - vertex-cut partitioning
  - vertex scopes
Good performance
Similar systems

Pregel, Giraph, GPS

GraphX → dataflow, on top of Spark

<table>
<thead>
<tr>
<th>System</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GraphLab</td>
<td>244</td>
</tr>
<tr>
<td>GraphX</td>
<td>251</td>
</tr>
<tr>
<td>Giraph</td>
<td>200</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>System</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GraphLab</td>
<td>249</td>
</tr>
<tr>
<td>GraphX</td>
<td>419</td>
</tr>
<tr>
<td>Giraph</td>
<td>596</td>
</tr>
</tbody>
</table>

twitter-2010, Connected Components
twitter-2010, PageRank, 20 iterations
Multilayer networks

social networks → multiplex networks
protein interaction networks
transportation
What is the best representation of multilayer networks in GraphLab/PowerGraph?

- flat graph with coloured edges
- different vertices for each layer

Test both implementations with PageRank
Experiment and Evaluation

Experiments run on an Amazon EC2 cluster
Evaluate performance measuring
  runtime
  memory consumption
  network IO
Project steps

Acquire multilayer datasets ← working on it
Setup GraphLab on Amazon EC2
Preprocess datasets to fit GraphLab input
Implement PageRank
Run multiple experiments on EC2
Analyse results
Write report 😊