Distributed Graph Processing - 2
Lecture 14

CSCI 4974/6971

20 Oct 2016
Today’s Biz

1. Reminders
2. Review
3. Assignment 4
4. Distributed Graph Processing
Reminders

- Assignment 3: solution out Friday
- Assignment 4: out Friday - due 3 November
  - Setting up and running on CCI clusters
- Project Update Presentation: In class November 3rd
- Office hours: Tuesday & Wednesday 14:00-16:00 Lally 317
  - Or email me for other availability
- Tentative class schedule:
  - Today: Distributed graph processing
  - Thursday: Distributed graph processing - more advanced
Today’s Biz

1. Reminders
2. Review
3. **Assignment 4**
4. Distributed Graph Processing
Assignment 4

1. Set up access to CCI clusters
2. Use distributed graph structure (Monday’s class) to run PageRank/BFS
3. Use partitioning methodology (Today’s class) to run PageRank/BFS
4. Examine strong scaling on fixed networks (comp/comm/idle)
5. Examine weak scaling on fixed networks (comp/comm/idle)
Today’s Biz

1. Reminders
2. Assignment 4
3. Review
4. Distributed Graph Processing
Quick Review

Distributed Graph Processing

1. Can’t store full graph on every node
2. Efficiently store local information - owned vertices / ghost vertices
   - Arrays for days - hashing is slow, not memory optimal
   - Relabel vertex identifiers
3. Vertex block, edge block, random, other partitioning strategies
### Quick Review

<table>
<thead>
<tr>
<th>Data</th>
<th>Size</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>n_global</td>
<td>1</td>
<td>Global vertex count</td>
</tr>
<tr>
<td>m_global</td>
<td>1</td>
<td>Global edge count</td>
</tr>
<tr>
<td>n_local</td>
<td>1</td>
<td>Task-local vertex count</td>
</tr>
<tr>
<td>n_ghost</td>
<td>1</td>
<td>Ghost vertex count</td>
</tr>
<tr>
<td>m_local_out</td>
<td>1</td>
<td>Task-local out-edges count</td>
</tr>
<tr>
<td>m_local_in</td>
<td>1</td>
<td>Task-local in-edges count</td>
</tr>
<tr>
<td>out_edges</td>
<td>m_out</td>
<td>Array of out-edges</td>
</tr>
<tr>
<td>out_offsets</td>
<td>n_loc</td>
<td>Start indices for local out-edges</td>
</tr>
<tr>
<td>in_edges</td>
<td>m_in</td>
<td>Array of in-edges</td>
</tr>
<tr>
<td>in_offsets</td>
<td>n_loc</td>
<td>Start indices for local in-edges</td>
</tr>
<tr>
<td>map</td>
<td>n_loc+n_gst</td>
<td>Global to local id hash table</td>
</tr>
<tr>
<td>local_unmap</td>
<td>n_loc</td>
<td>Array for local to global id conv.</td>
</tr>
<tr>
<td>ghost_unmap</td>
<td>n_gst</td>
<td>Array for local to global id conv.</td>
</tr>
<tr>
<td>tasks</td>
<td>n_gst</td>
<td>Array storing owner of ghost vertices</td>
</tr>
</tbody>
</table>
Today’s Biz

1. Reminders
2. Review
3. Assignment 4
4. Distributed Graph Processing
Distributed Processing
Blank code and data available on website
(Lecture 13)
www.cs.rpi.edu/~slotag/classes/FA16/index.html