Graph Twiddling in a MapReduce World
Jonathan Cohen

Adv. MapReduce Algorithms winter term 09/10
HPI
Winter presentation – Problem statement
Arvid Heise, Michael Leben
Use Case: Find Domains in Wikipedia

Graph Twiddling | Michael Leben, Arvid Heise | 11/17/2009
Use Case: Find Domains in Wikipedia

Graph Twiddling | Michael Leben, Arvid Heise | 11/17/2009
Use Case: Find Domains in Wikipedia
Map/Reduce Process for finding trusses

Original Graph → Simplify → Simplified Graph

Edges with Degrees → Augmentation

Enumerate Triangles

Find Trusses

Edges that form Triangles → Edges that are within trusses

Find Components

T1, T2
Simplification of Graph
Simplification of Graph

<table>
<thead>
<tr>
<th>Mapper</th>
<th>(hash, (GM, Opel))</th>
<th>(hash, (GM, Opel))</th>
<th>∅</th>
</tr>
</thead>
</table>

Mapper:

- {Opel, GM}
- {GM, Opel}
- {Audi, Audi}

Diagram:

- Arcandor
- Karstadt
- Quelle
- Opel
- Audi
- Car
- Short-time working

Graph Twiddling | Michael Leben, Arvid Heise | 11/17/2009
Simplification of Graph

Mapper

| {Opel, GM}   | (hash, (GM, Opel)) |
| {GM, Opel}   | (hash, (GM, Opel)) |
| {Audi, Audi} | ∅                  |

Reducer

| (hash, (GM, Opel)), (hash, (GM, Opel)), ... | (GM, Opel) |
Simplification of Graph
Augmenting Edges with Degrees
Augmenting Edges with Degrees

![Graph visualisation](image-url)
Ins=5, GM=3
Augmenting Edges with Degrees

Mapper 1

| (GM, Opel) | (Opel, (GM, Opel)), (GM, (GM, Opel)) |

Ins=5, GM=3

Arcandor

Karstadt

Quelle

Opel

Audi

Short-time working

GM

Insolvency

11/17/2009
Augmenting Edges with Degrees

Ins=5, GM=3

Mapper1

| (GM, Opel) | (Opel, (GM, Opel)), (GM, (GM, Opel)) |

Reducer1

| (GM, (GM, Opel)), (GM, (GM, Ins)), (GM, (Car, GM)) | ((GM, Opel), GM=3, Opel=), ((GM, Ins), GM=3, Ins=), ((Car, GM), Car=, GM=3) |
Augmenting Edges with Degrees

<table>
<thead>
<tr>
<th>Mapper1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(GM, Opel)</td>
<td>(Opel, (GM, Opel)), (GM, (GM, Opel))</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reducer1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(GM, (GM, Opel)), (GM, (GM, Ins)), (GM, (Car, GM))</td>
<td>(((GM, Opel), GM=3, Opel=), ((GM, Ins), GM=3, Ins=), ((Car, GM), Car=, GM=3))</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reducer2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>((GM, Ins), GM=3, Ins=), ((GM, Ins), GM=, Ins=5)</td>
<td>(((GM, Ins), GM=3, Ins=5))</td>
</tr>
</tbody>
</table>
In Detail: Finding Triangles  1|5

Input

Graph

Twiddling

### Input

<table>
<thead>
<tr>
<th>Relation</th>
<th>Distance (Insolvency)</th>
<th>Distance (Arcandor)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Insolvency, Arcandor)</td>
<td>D(Insolvency) = 5</td>
<td>D(Arcandor) = 3</td>
</tr>
<tr>
<td>(Arcandor, Quelle)</td>
<td>D(Arcandor) = 3</td>
<td>D(Quelle) = 3</td>
</tr>
<tr>
<td>(Insolvency, GM)</td>
<td>D(Insolvency) = 5</td>
<td>D(GM) = 3</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
In Detail: Finding Triangles  2|5

Idea

Triad

Apex

Endpoints

Insolvency

5

Arcandor

3

Karstadt

3

Quelle

3
### Output of first Mapper:
**Edges to Apex**

<table>
<thead>
<tr>
<th>Key (Apex)</th>
<th>Edges</th>
</tr>
</thead>
</table>
| **Arcandor** | (Arcandor, Insolvency)  
                    (Arcandor, Karstadt)  
                    (Arcandor, Quelle)  |
| **Quelle**  | (Insolvency, Quelle)                       |
| **Karstadt** | (Insolvency, Karstadt)  
                    (Karstadt, Quelle)  |

[Diagram of a graph with nodes labeled Insolvency, Arcandor, Karstadt, and Quelle, connected by edges, showing the relationships between the companies.]
1. Reduce: Form Triads

**Output of first Reducer:**
Emit Triads, keyed by endpoints

<table>
<thead>
<tr>
<th>Key</th>
<th>Triad</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Insolvency, Karstadt</strong></td>
<td>(Arcandor, Insolvency) (Arcandor, Karstadt)</td>
</tr>
<tr>
<td>Karstadt, Quelle</td>
<td>(Arcandor, Karstadt) (Arcandor, Quelle)</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Graph Twiddling | Michael Leben, Arvid Heise | 11/17/2009
### Output of second Mapper:
Map edges and triads to their endpoints

<table>
<thead>
<tr>
<th>Key</th>
<th>Triad or Edge</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Insolvency, Karstadt</strong></td>
<td>Arcandor, Insolvency, Arcandor, Karstadt</td>
</tr>
<tr>
<td><strong>Insolvency, Karstadt</strong></td>
<td>Insolvency, Karstadt</td>
</tr>
<tr>
<td><strong>...</strong></td>
<td></td>
</tr>
</tbody>
</table>

Graph Twiddling | Michael Leben, Arvid Heise | 11/17/2009
In Detail: Finding Triangles  5|5

2. Phase: Close Triads with Edges

Output of second Mapper:
- Map edges and triads to their endpoints
- Insolvency, Karstadt Arcandor

Output of second Reducer:
- Emit Triangles
- Merge Edges and Triads to Triangles

<table>
<thead>
<tr>
<th>Key</th>
<th>Triangle</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>Arcandor, Insolvency Arcandor, Karstadt Insolvency, Karstadt</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Graph Twiddling | Michael Leben, Arvid Heise | 11/17/2009
Process Overview

Original Graph → Simplify → Simplified Graph

Edges with Degrees → Augmentation → Edges that form Triangles

Enumerate Triangles

Find Trusses

Edges that are within trusses

Find Components

T1, T2
Finding Trusses

Graph Twiddling | Michael Leben, Arvid Heise | 11/17/2009
Finding Trusses

Graph Twiddling | Michael Leben, Arvid Heise | 11/17/2009
Finding Trusses

Graph Twiddling | Michael Leben, Arvid Heise | 11/17/2009
Mapper

| (GM, Ins, Opel) | ((GM, Opel), 1), ((GM, Ins), 1), ((Ins, Opel), 1) |

Finding Trusses

Graph Twiddling | Michael Leben, Arvid Heise | 11/17/2009
Finding Trusses

Mapper

| (GM, Ins, Opel) | ((GM, Opel), 1), ((GM, Ins), 1), ((Ins, Opel), 1) |

Reducer

| ((GM, Ins), 1) | ∅ |
| ((GM, Opel), 1), ((GM, Opel), 1) | (GM, Opel) |
Finding Trusses #2

Graph:
- Insolvency
- Arcandor
- Karstadt
- Quelle
- GM
- Car
- Opel
- Audi

Legend:
- Short-time working
Finding Trusses #2

Graph Twiddling | Michael Leben, Arvid Heise | 11/17/2009

- Insolvency
- Arcandor
- Karstadt
- Quelle
- GM
- Opel
- Car
- Audi
- Short-time working

0
Finding Trusses #2

Insolvency

Arcandor

Karstadt

Quelle

Short-time working

Car

Opel

Audi
Finding Components

- Expands zones from all nodes simultaneously
- Merges incrementally zones in two steps (3 MapReduce Jobs)
- Expands zones from all nodes simultaneously
- Merges incrementally zones in two steps (3 MapReduce Jobs)
Expands zones from all nodes simultaneously
Merges incrementally zones in two steps (3 MapReduce Jobs)
Expands zones from all nodes simultaneously

Merges incrementally zones in two steps (3 MapReduce Jobs)
Outlook

- Implement 5 algorithms
  - Simplification, augmentation of edges with degrees ✓
  - Enumeration of triangles, trusses, and components

- Evaluation
  - Resource usage / scalability
    - Amount of data might be bottleneck
  - Is it applicable for all graphs?
    - DBpedia and sub graphs as test data
  - Description of algorithms