Paramount task of YouLighter:
- Study evolution of YouTube infrastructure
- Highlight change in YouTube infrastructure

Motivation:
- It generates 20+% of world wide traffic
- YouTube has a massive distributed infrastructure that is almost unknown
- It uses several thousands of caches (single server) grouped into Hundreds of edge-nodes
- This infrastructure suddenly evolve

Monitoring the single cache is not effective
- Load distribution changes very frequently
- The rank of most used caches changes deeply everyday!

Dataset

<table>
<thead>
<tr>
<th>Probe</th>
<th>Period</th>
<th>Volume</th>
<th># Unique Videos</th>
<th>Caches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe 1 Italy</td>
<td>01/04/2013 - 28/02/2014</td>
<td>138.7 TB</td>
<td>2,892,452</td>
<td>8,664</td>
</tr>
<tr>
<td>Probe 1 Italy</td>
<td>01/04/2013 - 28/02/2014</td>
<td>152.9 TB</td>
<td>2,848,625</td>
<td>8,899</td>
</tr>
<tr>
<td>Probe 2 Italy</td>
<td>01/04/2013 - 28/02/2014</td>
<td>134.8 TB</td>
<td>2,711,179</td>
<td>9,028</td>
</tr>
<tr>
<td>Probe 3 Poland</td>
<td>01/03/2014 - 17/07/2014</td>
<td>48.3 TB</td>
<td>305,802</td>
<td>3,755</td>
</tr>
</tbody>
</table>

Monitoring the single cache is not effective
- Load distribution changes very frequently
- The rank of most used caches changes deeply everyday!

Constellation Distance CD
1. Summarize each cluster in a single point called star $\hat{C}_n$
2. Astra Distance
   - For each star in $\hat{C}_n$ compute all distances to stars in $\hat{C}_{n+1}$ and the min
   - Repeat in the opposite direction
3. Constellation Distance
   - Sum all Astra Distances

Results

Conclusion: YouLighter shows to be effective at detecting changes in YouTube’s CDN infrastructure relying on DBSCAN clustering algorithm and the novel notion of Constellation distance.