

Visual analysis of massive web session data

<http://ieeexplore.ieee.org/xpls/icp.jsp?arnumber=6378977>

这篇文章实现了对于比较大规模的信息数据在分布式系统上的可视化，与我们之前的设想十分相近。

<http://www.trifacta.com/>

这家公司的产品实现了将 HDFS 上的数据通过可视化的手段进行浏览，作为数据清洗的工具。同时也加入了诸如 Machine Learning 之类的元素。

综上，之前的那个思路可能并没有太大贡献。

经过和陈老师的讨论，重新将注意转到了利用 Spark (GraphX) 对大图进行可视化。

与方舟师兄讨论了这个方案的可行性。

我们认为可行的流程是这样的：

对于 Overlay，可以在集群上对图进行聚类，将聚类结果交给前端可视化。

对于 N 度节点，可以尝试将现有的力引导图并行算法放到集群上计算。将计算结果传给前端来显示。

当点和边的数量太大时，使用 SVG 不太显示，可能需要用 WebGL，或者将整个 canvas 从后端传上来。

GraphX 的文档：

<http://spark.apache.org/docs/latest/graphx-programming-guide.html>

现在已有一些将大图并行的方案：

A Scalable Parallel Force-Directed Graph Layout Algorithm

不过是基于 GPU 的，不知是否有可能将其思路转移到集群计算上。

因为大图可视化这个问题的研究者可能很多，下一步的主要工作是查找 Related Work。