Big Data Visualization Using Cloudberry

Er Zhuo, Yi Zhao, Qiandong Tang
Department of Computer Science, University of California, Irvine
{erzhuo1026, joy11612917, qiandongt}@gmail.com

UCI Samueli School of Engineering

Background

- **Cloudberry** is an open source platform for big data visualization.
- Based on Cloudberry, **Twittermap** is an application for **interactive analytics** and **visualization** of more than **1.6 billion** tweets, which are rich with **temporal, spatial, and textual** attributes.
- As an extremely popular **social network**, not only can Twitter support isolated tweets, it can also let users **interacts with each other** through “retweeting”.
- Therefore, to **effectively visualize tweets as a graph** is highly desired.

Introduction

- Goal: **Interactively** show an **uncluttered** graph of reply tweets containing a certain keyword to users.
- Uncluttered: Vertex Clustering and Edge Bundling.
- Interactive: Incrementally query with time slices.

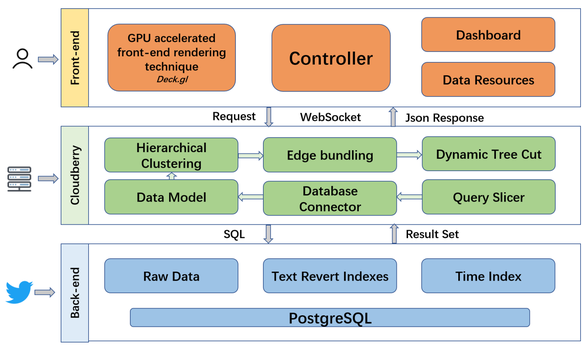


Fig. 1 System workflow

Methodology

- **Incremental Query**: Slice the database query by time dimension.
- **Vertex Clustering**: Use the incrementalized version of *Hierarchical Greedy Clustering* to cluster vertices. The hierarchical structure of clusters is stored in the middleware. Different number of clusters can be displayed under different zoom levels.
- **Edge Bundling**: Use *Force Directed Edge Bundling* [1] algorithm to reduce the cluttered edges by applying forces according to physical formulas.
- **Tree Cut Algorithm**: Dynamically find a smaller set of clusters in the hierarchical structure that minimizes the messy orientations of edges as well as preserves the geographical information.

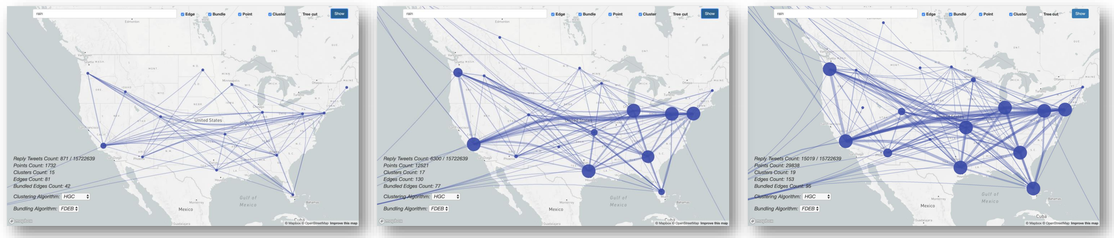
Conclusion

- Our system is a general middleware solution without changing the underlying database system.
- Our results on over 1.6 billion tweets showed that our system and its techniques can offer better user experience by incrementalizing the whole pipeline as well as reduce the clutters of the graph visualization.

Reference

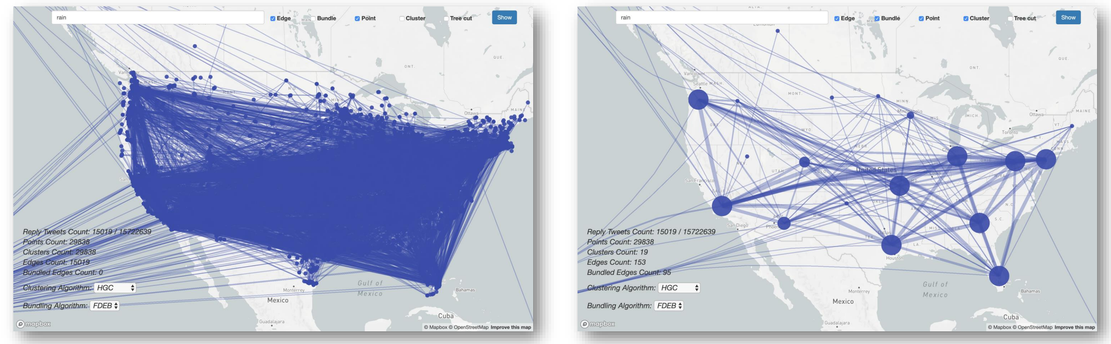
[1] Holten, D., & Van Wijk, J. J. (2009, June). Force-directed edge bundling for graph visualization. In Computer graphics forum (Vol. 28, No. 3, pp. 983-990). Oxford, UK: Blackwell Publishing Ltd.

Results



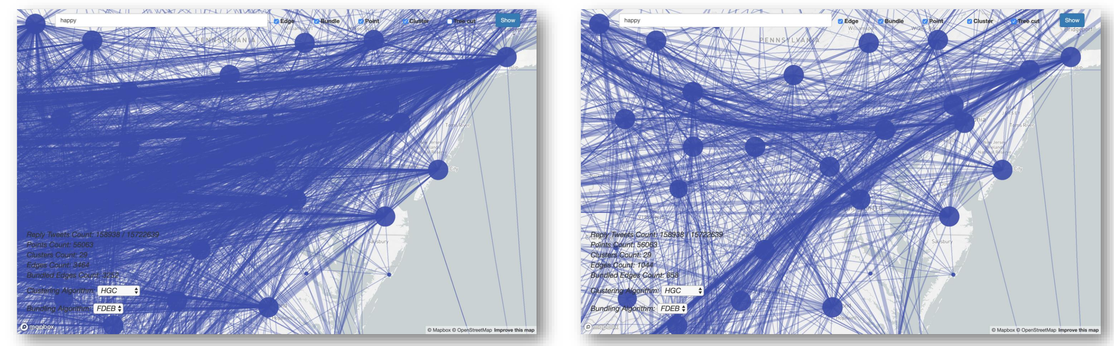
(a) Progressive step 1 (b) Progressive step 2 (c) Progressive step 3

Fig. 2 Progressive Visualization Process



(a) Original Graph (b) Simplified Graph

Fig. 3 Comparison between original graph and simplified graph



(a) Zoom in (b) Zoom in with tree cut

Fig. 4 Comparison between graph with and without tree cut