Large-scale Geometric Graph Visualization

Authors: Er Zhuo, Yi Zhao, Qiandong Tang   Advisors: Chen Li, Shuang Zhao, Sadeem Alsudais, Qiushi Bai
Department of Computer Science, University of California, Irvine
erzhuo1026@gmail.com, joy11612917@gmail.com, qiandongt@gmail.com
chenli@gmail.com, shz@ics.uci.edu, sadeem.alsudais@gmail.com, baiqiushi@gmail.com
https://cloudberry.ics.uci.edu

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 **interact 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.
- Uncluttered: Vertex Clustering and Edge Bundling.
- Interactive: Incrementally query by slicing time.

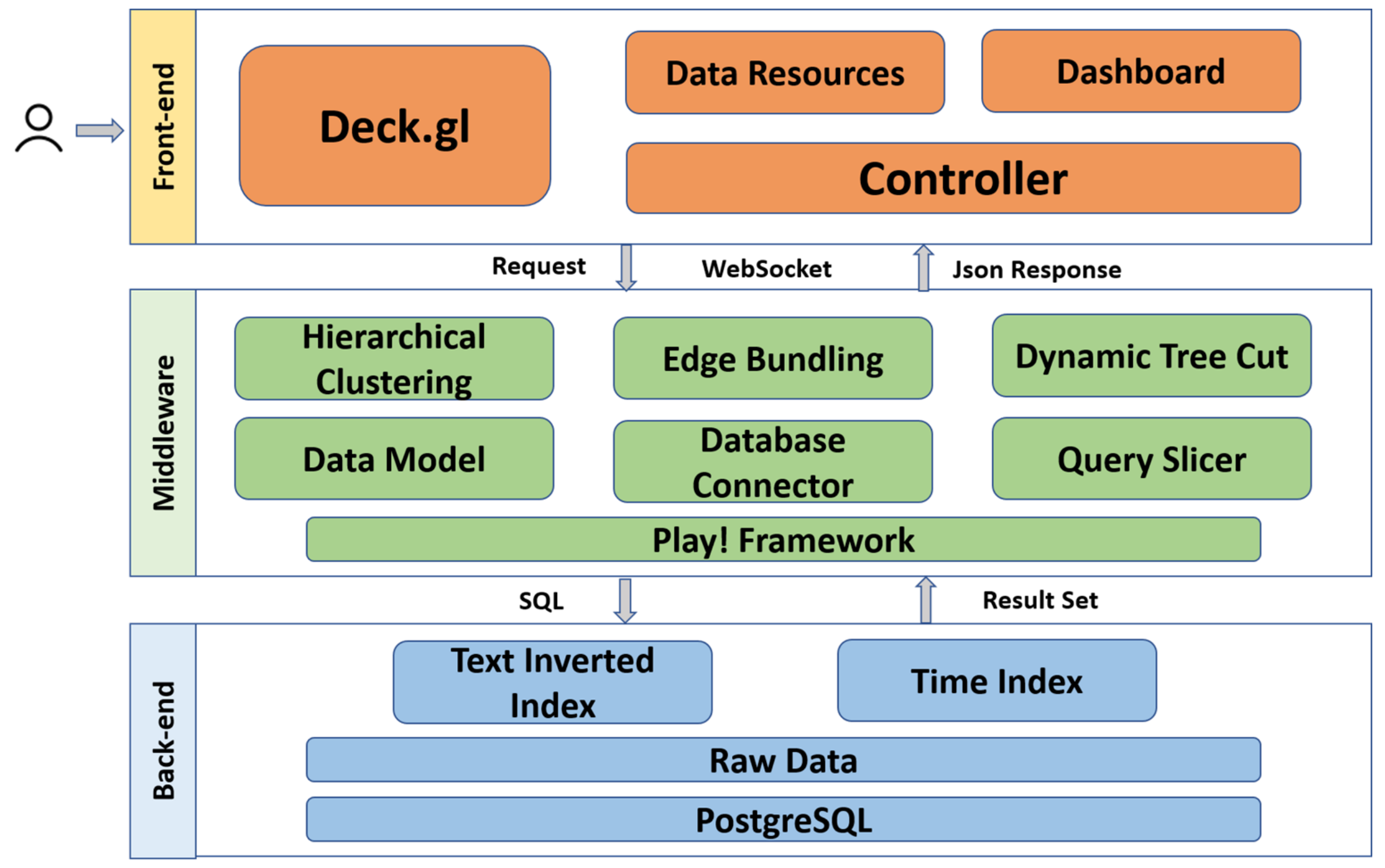


Fig. 1 System Architecture

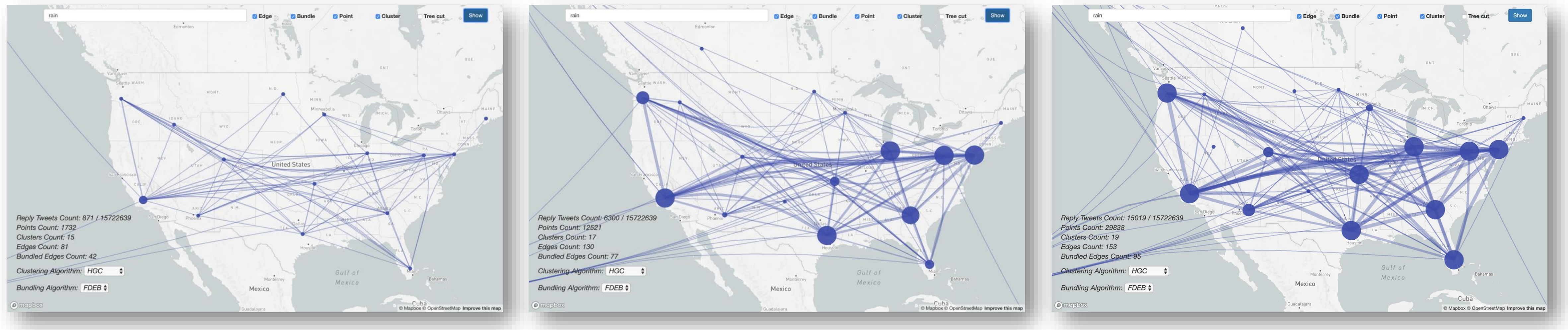
Methodology

- **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* algorithm to reduce the cluttered edges by applying force according to physical formulas.
- **Tree Cut Algorithm**: Dynamically find a smaller set of clusters in the hierarchical structure that minimize the messy orientations of edges as well as preserve the geometric information.

Conclusion

- Our solution is a database-driven system for visualizing large-scale geometric graphs.
- Our results on over 15 million tweets showed that our system and its techniques can offer better user experience by incrementalizing the whole pipeline as well as reducing the clutters of the graph visualization.

Results



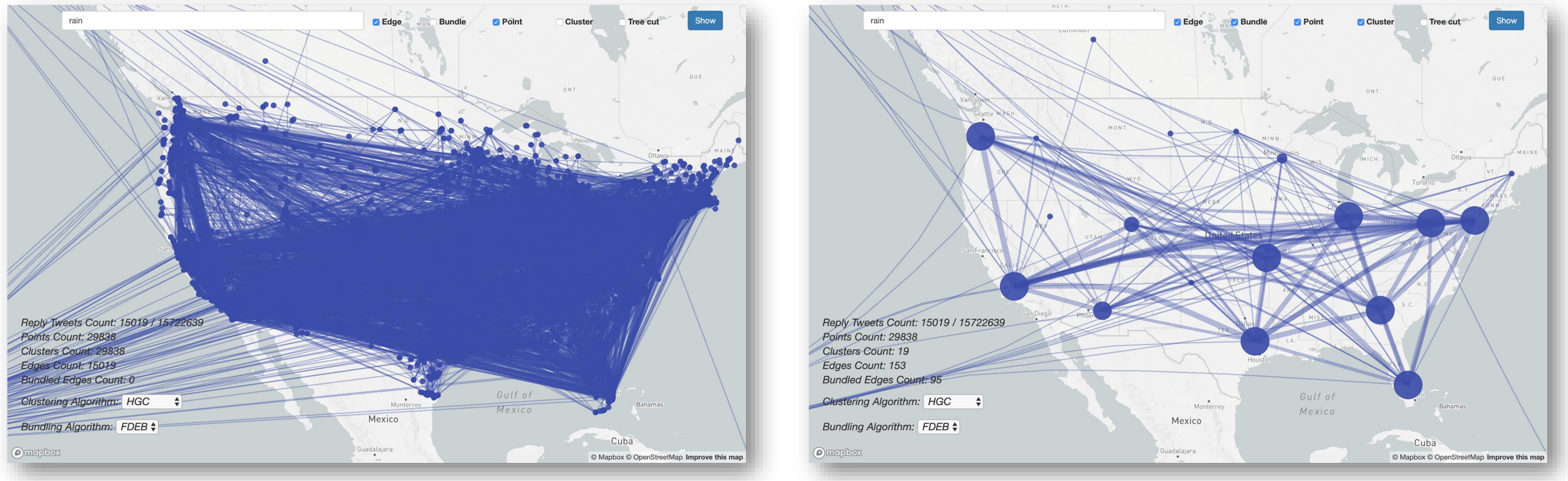
(a) Reply Tweets Count: 871 / 15722639
Points Count: 1732
Clusters Count: 15
Edges Count: 81
Bundled Edges Count: 42
Clustering Algorithm: HGC
Bundling Algorithm: FDEB

(b) Reply Tweets Count: 6300 / 15722639
Points Count: 12521
Clusters Count: 17
Edges Count: 130
Bundled Edges Count: 77
Clustering Algorithm: HGC
Bundling Algorithm: FDEB

(c) Reply Tweets Count: 15019 / 15722639
Points Count: 29838
Clusters Count: 19
Edges Count: 153
Bundled Edges Count: 95
Clustering Algorithm: HGC
Bundling Algorithm: FDEB

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

Fig. 2 Progressive visualization process

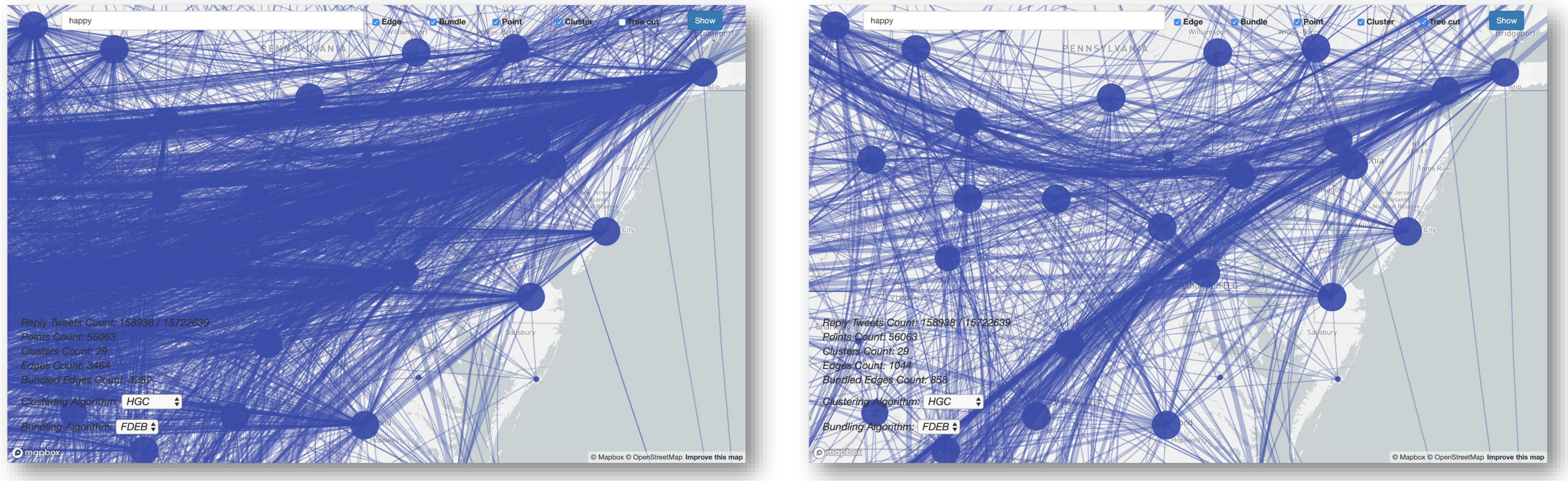


(a) Reply Tweets Count: 15019 / 15722639
Points Count: 29838
Clusters Count: 29838
Edges Count: 15019
Bundled Edges Count: 0
Clustering Algorithm: HGC
Bundling Algorithm: FDEB

(b) Reply Tweets Count: 15019 / 15722639
Points Count: 29838
Clusters Count: 19
Edges Count: 153
Bundled Edges Count: 95
Clustering Algorithm: HGC
Bundling Algorithm: FDEB

(a) Original Graph   (b) Simplified Graph

Fig. 3 Comparison between the original graph and the simplified graph



(a) Reply Tweets Count: 158938 / 15722639
Points Count: 56063
Clusters Count: 29
Edges Count: 3464
Bundled Edges Count: 3252
Clustering Algorithm: HGC
Bundling Algorithm: FDEB

(b) Reply Tweets Count: 158938 / 15722639
Points Count: 56063
Clusters Count: 29
Edges Count: 1044
Bundled Edges Count: 858
Clustering Algorithm: HGC
Bundling Algorithm: FDEB

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

Fig. 4 Comparison between graphs with and without tree cut