

Visualizing Connection Pathways in Large Graphs

Polo Chau
Carnegie Mellon

Leman Akoglu
Carnegie Mellon

Jilles Vreeken
University of Antwerp

Hanghang Tong
IBM T.J. Watson Research

Christos Faloutsos
Carnegie Mellon

TourViz is an interactive system that reveals and visualizes connections among nodes of interest.

The Problem

How to make sense of *marked nodes* in a large graph (e.g., anomalies) ?

How are they connected?
Are they close by or far apart?
Are there simple paths that connect them?
Can we group them?

Algorithm

Our algorithm finds **simple** paths between marked nodes, and group them, based on the **Minimum Description Length** principle, so that each path needs few bits to describe, e.g., avoid high-degree nodes, unless need to visit many of its spokes.

This is an NP-hard problem. Our fast algorithm is described in [1].

Interactive Visualization

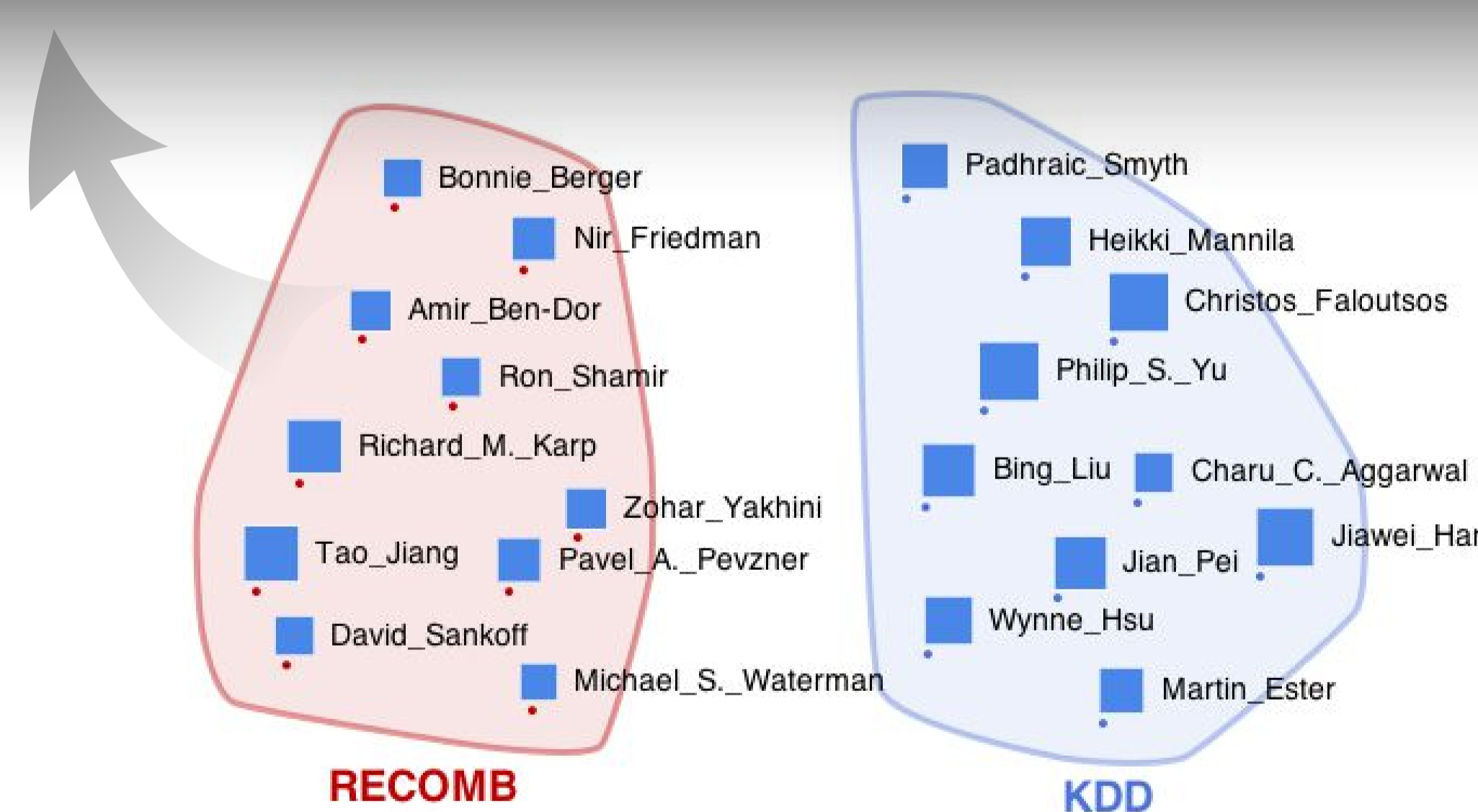
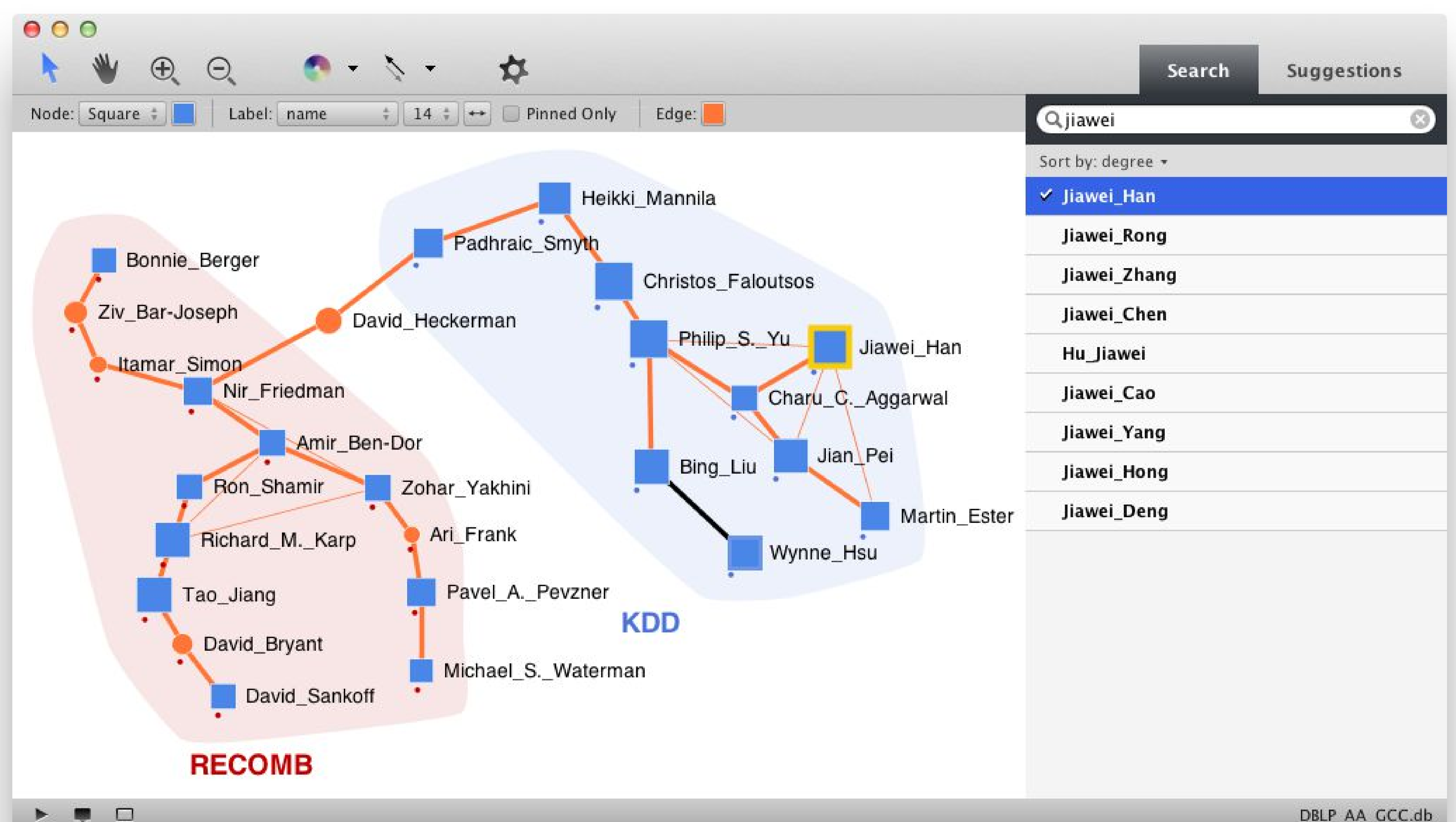
1 Search. Select.
Find nodes and drag them into the view.

2 Mark nodes. Go.
Turn your nodes of interest into **squares**.
TourViz will find simple paths among them.

You can also group nodes visually.

3 Visualize. Interact.
TourViz visualizes paths among marked nodes.

You can interact with them: add or delete nodes, mark or unmark them, see their neighbors, and more.



TourViz showing connection pathways among authors from DBLP coauthorship graph (300K nodes, 1M edges).

- **Blue square:** nodes of interest
- **Orange circle:** connectors
- **Thick orange edge:** simple path found by TourViz

[1] L. Akoglu, J. Vreeken, H. Tong, D. H. Chau, and C. Faloutsos. Islands and bridges: Making sense of marked nodes in large graphs. Technical Report CMU-CS-12-124, Carnegie Mellon University, 2012.

Visualization implemented in Java, using the JUNG library. Algorithm written in Matlab 7.10.

Research was sponsored by the National Science Foundation under Grant No. IIS1017415 and the Army Research Laboratory under Cooperative Agreement No. W911NF-09-2-0053. It is continuing through participation in the Anomaly Detection at Multiple Scales (ADAMS) program sponsored by the U.S. Defense Advanced Research Projects Agency (DARPA) under Agreements No. W911NF-11-C-0200 and W911NF-11-C-0088. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory, of the National Science Foundation, of the U.S. Government, or any other funding parties. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on. This work is also partially supported by an IBM Faculty Award. Jilles Vreeken is supported by a Post-Doctoral Fellowship of the Research Foundation – Flanders (FWO).