

# OPAvion

## Mining & Visualization in Large Graphs

Leman Akoglu  
lakoglu@cs.cmu.edu

Polo Chau  
dchau@cs.cmu.edu

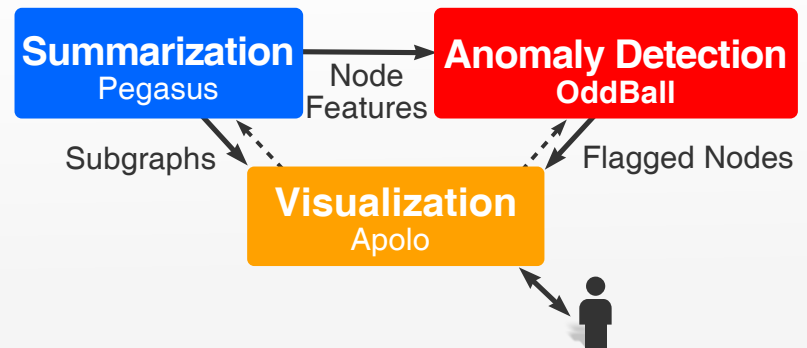
U Kang  
ukang@cs.cmu.edu

Danai Koutra  
danai@cs.cmu.edu

Christos Faloutsos  
christos@cs.cmu.edu

Given a billion-node graph, like a who-follows-whom Twitter graph, how do we summarize its patterns, spot anomalies, and visualize it? Meet **OPAvion**, a scalable graph mining system, with three modules:

- **Summarization** (Pegasus)
- **Anomaly Detection** (OddBall)
- **Visualization** (Apolo)

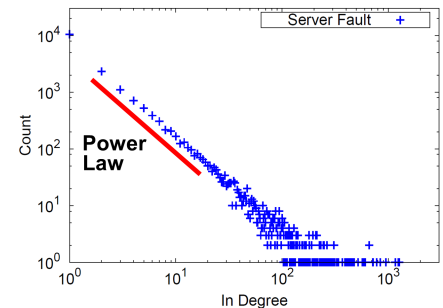


## Summarization

Pegasus (award-winning)



Operates off-line on billion-node, disk-resident graphs and computes statistics, like PageRank, connected components, degree distribution, triangles, etc.

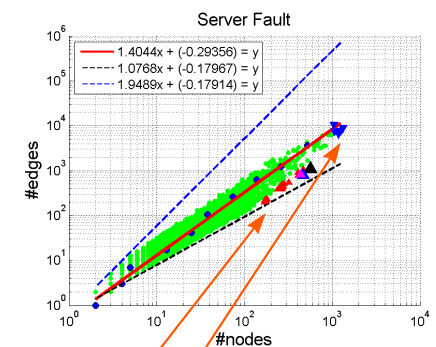


## Anomaly Detection

OddBall (award-winning)



Uses graph statistics to mine patterns and spot anomalies, such as nodes with many contacts but few interactions with them (possibly telemarketers).



## Visualization

Apolo

Allows users to incrementally explore the graph, starting with their chosen nodes or the flagged anomalous nodes; then users can expand to the nodes' vicinities, label them into categories, and interactively navigate to other relevant parts of the graph.

