Carnegie Mellon

15-826: Multimedia Databases and Data Mining

Project lecture #2: Anomaly detection in large graphs

Christos Faloutsos

Carnegie Mellon

Roadmap

- Introduction – Motivation
- Patterns in graphs
  - Static graphs
  - <etc>
- Tools:
  - OddBall
  - CatchSync
- Conclusions

Carnegie Mellon

Problem

- Tools, for anomaly detection in graphs?

Carnegie Mellon

Conclusions

- Tools, for anomaly detection in graphs?
- MANY – two of them:
  - OddBall (features of ego-nets)
  - CatchSync (in-degree vs 'authoritativeness', etc)

Carnegie Mellon

OddBall: Spotting anoMalies in Weighted Graphs


Leman Akoglu, Mary McGlohon, Christos Faloutsos

Carnegie Mellon University

School of Computer Science

PAKDD 2010, Hyderabad, India

Carnegie Mellon

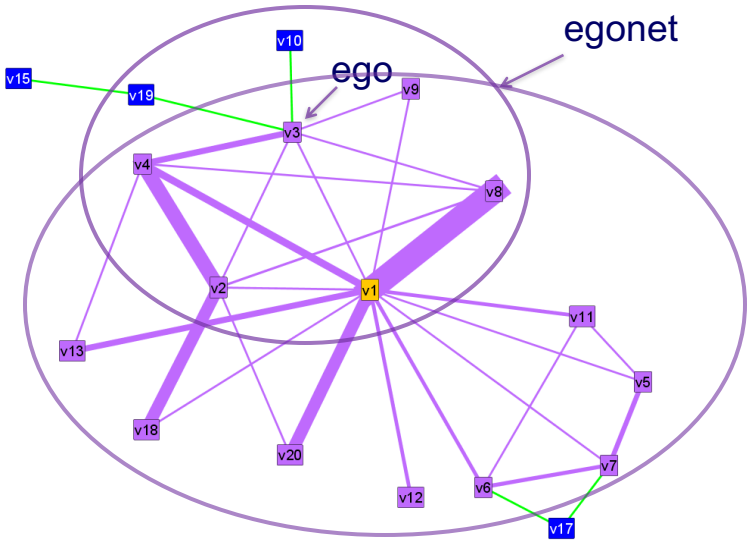
Main idea

For each node,

- extract 'ego-net' (=1-step-away neighbors)
- Extract features (#edges, total weight, etc etc)
- Compare with the rest of the population

Carnegie Mellon

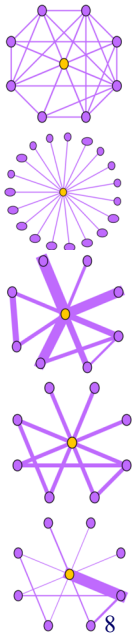
What is an egonet?



Carnegie Mellon

Selected Features

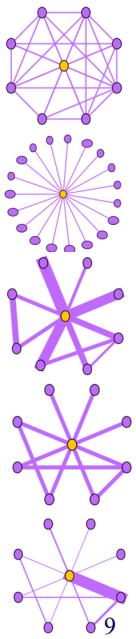
Q: which features?
Q': how many possible features exist?



Carnegie Mellon

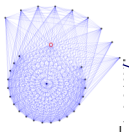
Selected Features

- N_i: number of neighbors (degree) of ego i
- E_i: number of edges in egonet i
- W_i: total weight of egonet i
- $\lambda_{w,i}$: principal eigenvalue of the weighted adjacency matrix of egonet I



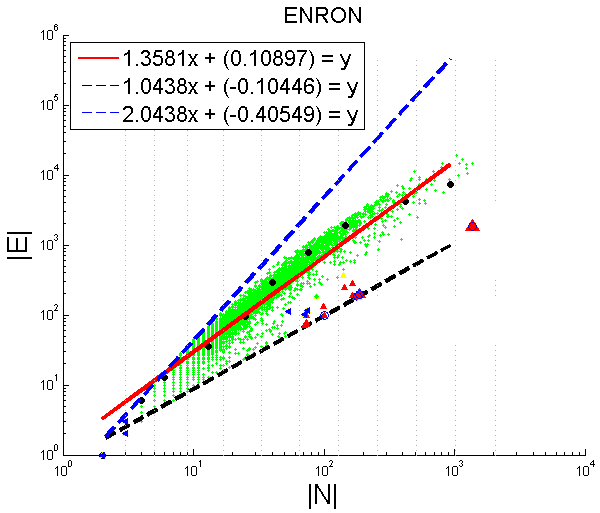
Carnegie Mellon

Near-Clique/Star



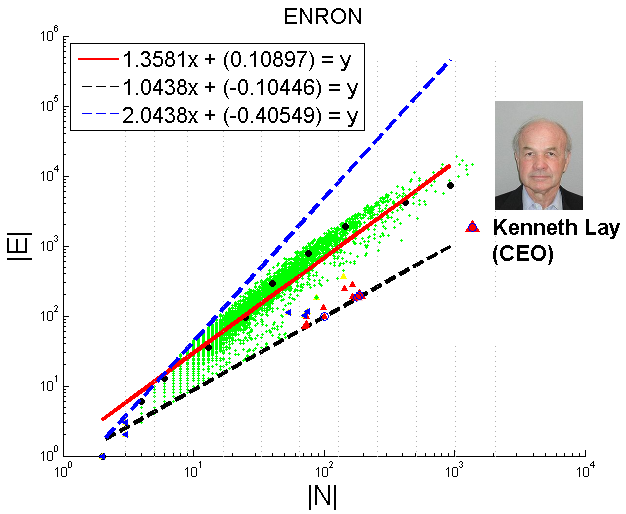
Carnegie Mellon

Near-Clique/Star



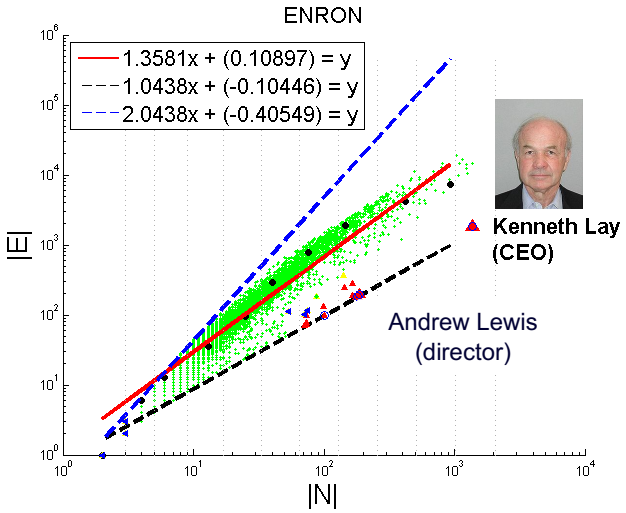
Carnegie Mellon

Near-Clique/Star



Carnegie Mellon

Near-Clique/Star



Carnegie Mellon

Roadmap

- Introduction – Motivation
- Patterns in graphs
  - Static graphs
  - <etc>
- Tools:
  - OddBall
  - CatchSync
- Conclusions

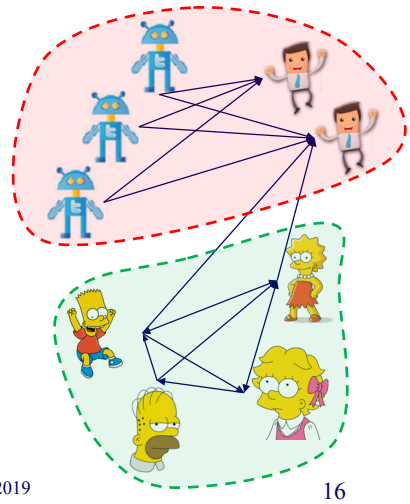
Carnegie Mellon

Catchsync: catch synchronized behavior in large directed graphs

Meng Jiang, Peng Cui, Alex Beutel,
Christos Faloutsos and Shiqiang Yang
KDD, August 26, 2014 – NYC, USA

Carnegie Mellon

Fraud Detection: Graph Analysis Problem

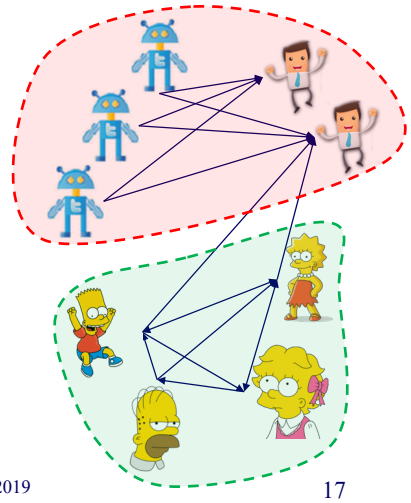


Carnegie Mellon

Fraud Detection: Graph Analysis Problem

[www.buyfollowz.org]

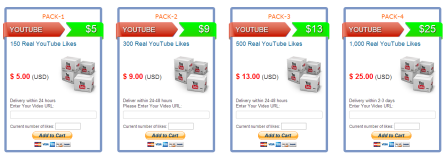
[buymorelikes.com]



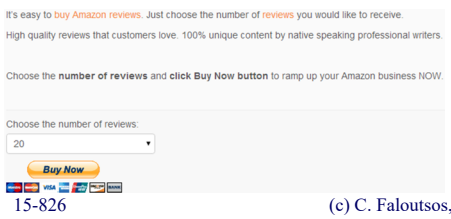
Carnegie Mellon

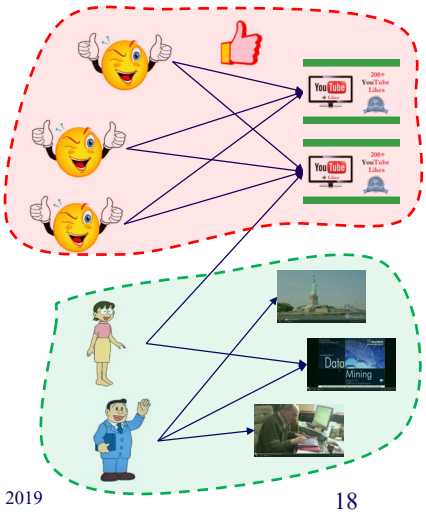
Fraud Detection: Graph Analysis Problem

[buycheaplikes.com]



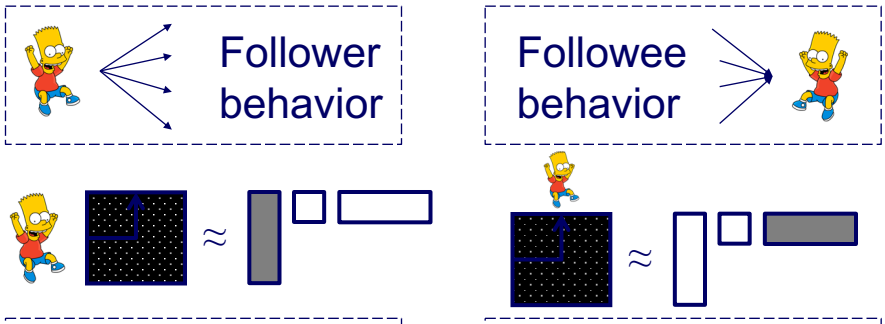
[reviewsteria.com]





Carnegie Mellon

Behavior-based Features



Out-degree
1st left singular vector
(Hubness)
…

In-degree
1st right singular vector
(Authoritativeness)
…

Carnegie Mellon

Hubs/Authorities

Jon M. Kleinberg: *Authoritative Sources in a Hyperlinked Environment*. SODA 1998: 668-677

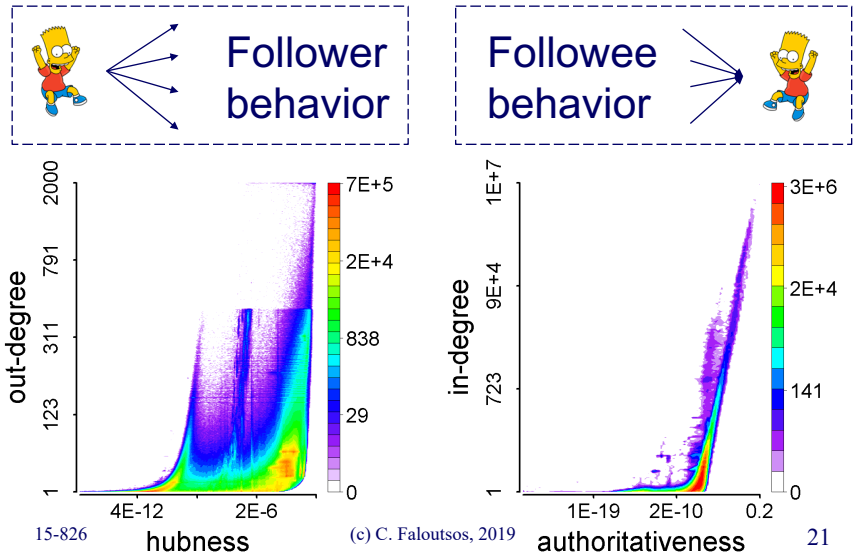
(Later, in huge detail…)

```
import networkx as nx
G=nx.path_graph(4)
h,a=nx.hits(G)

print h
print a
```

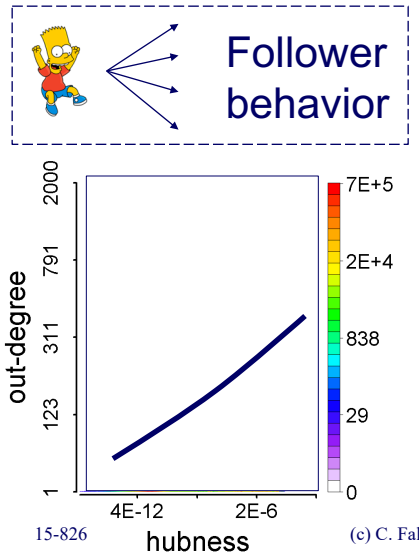
Carnegie Mellon

Behavior-based Feature Space



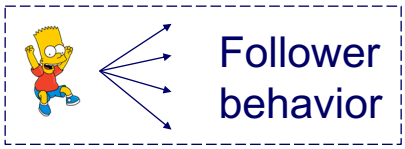
Carnegie Mellon

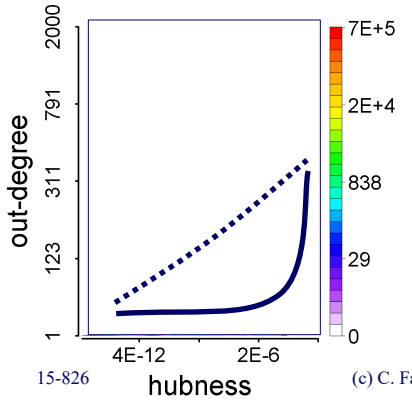
Behavior-based Feature Space



Carnegie Mellon

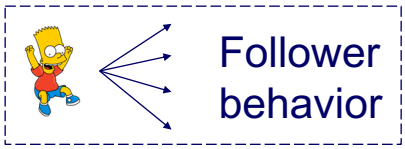
Behavior-based Feature Space

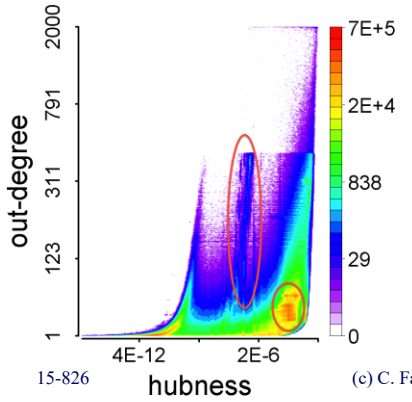




Carnegie Mellon

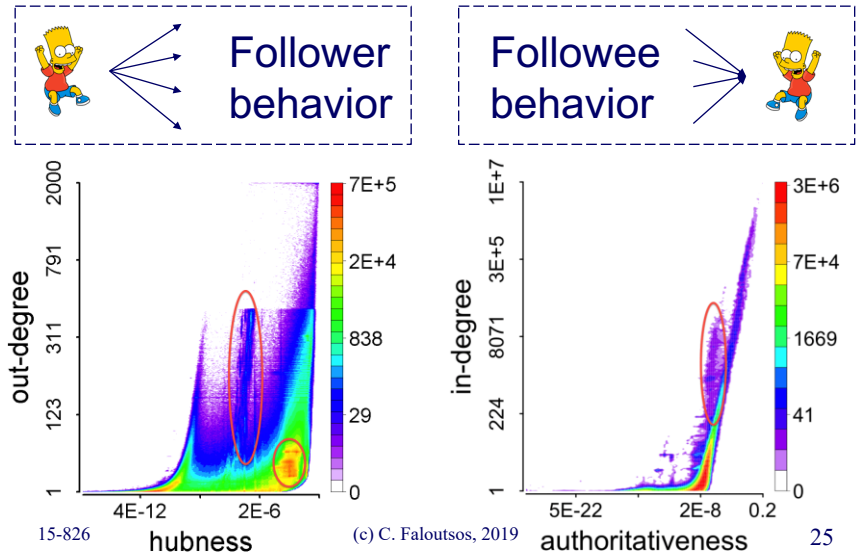
Before CatchSync





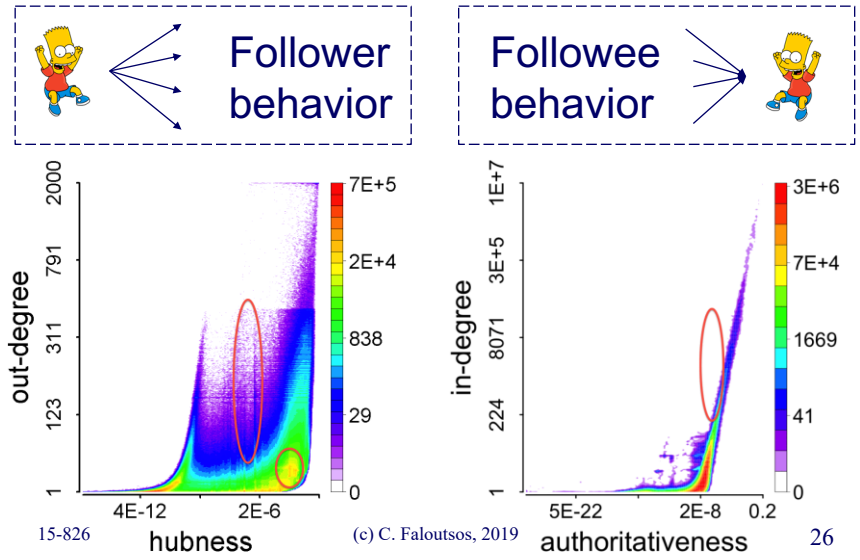
Carnegie Mellon

Before CatchSync



Carnegie Mellon

After CatchSync



Carnegie Mellon

Conclusions

- Tools, for anomaly detection in graphs?
- MANY – two of them:
  - OddBall (features of ego-nets)
  - CatchSync (in-degree vs 'authoritativeness', etc)