

CarnegieMellon

15-826: Multimedia Databases and Data Mining

Project lecture #2: Anomaly
detection in large graphs

Christos Faloutsos


CarnegieMellon

Roadmap

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



CarnegieMellon




Problem

- Tools, for anomaly detection in graphs?

15-826 Copyright: C. Faloutsos (2019) 3

CarnegieMellon



Conclusions

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

15-826 Copyright: C. Faloutsos (2019) 4

CarnegieMellon

OddBall: Spotting anomalies in Weighted Graphs



Leman Akoglu, Mary McGlohon, Christos
Faloutsos

*Carnegie Mellon University
School of Computer Science*

PAKDD 2010, Hyderabad, India

CarnegieMellon

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

CarnegieMellon

What is an egonet?

15-826 (c) C. Faloutsos, 2019 7

CarnegieMellon

Selected Features

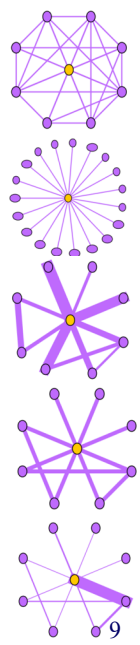
Q: which features?
Q': how many possible features exist?

15-826 (c) C. Faloutsos, 2019

CarnegieMellon

Selected Features

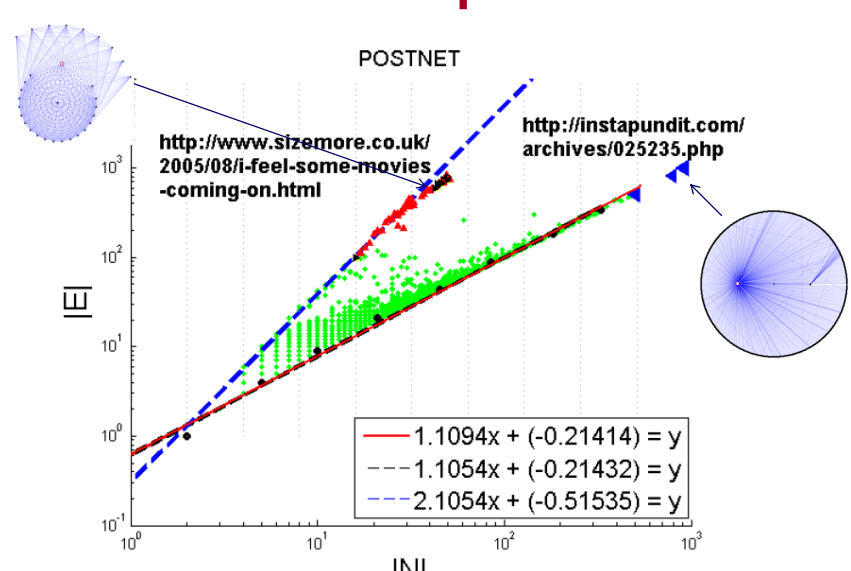
- N_i : number of neighbors (degree) of ego i
- E_i : number of edges in egonet i
- W_i : total weight of egonet i
- $\lambda_{w,i}$: principal eigenvalue of the **weighted** adjacency matrix of egonet I



15-826 (c) C. Faloutsos, 2019

CarnegieMellon

Near-Clique/Star



POSTNET

<http://www.sizemore.co.uk/2005/08/i-feel-some-movies-coming-on.html>

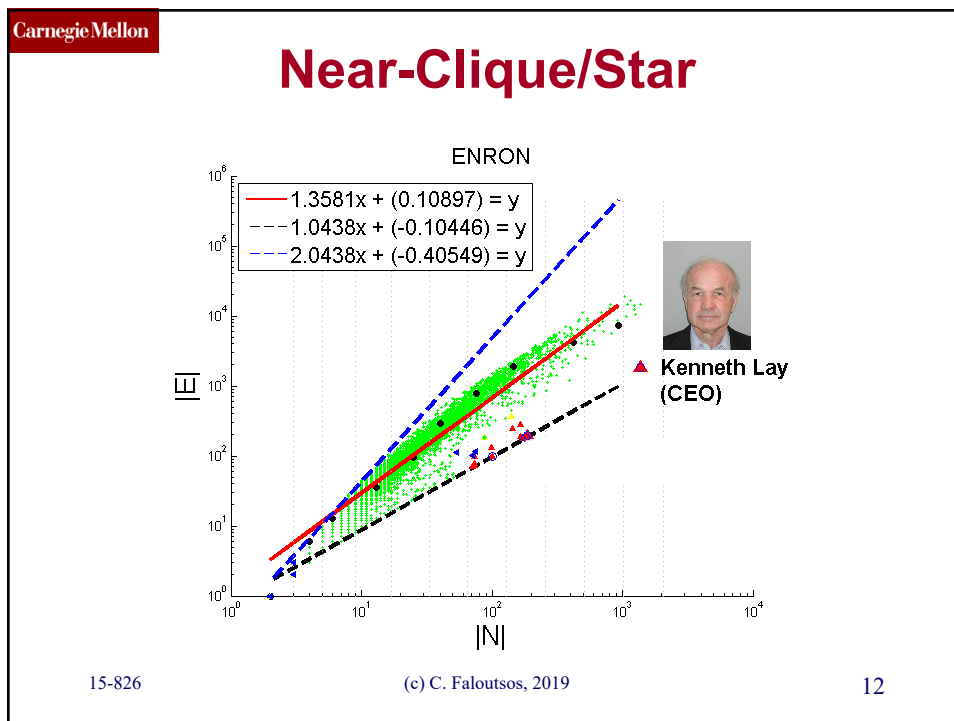
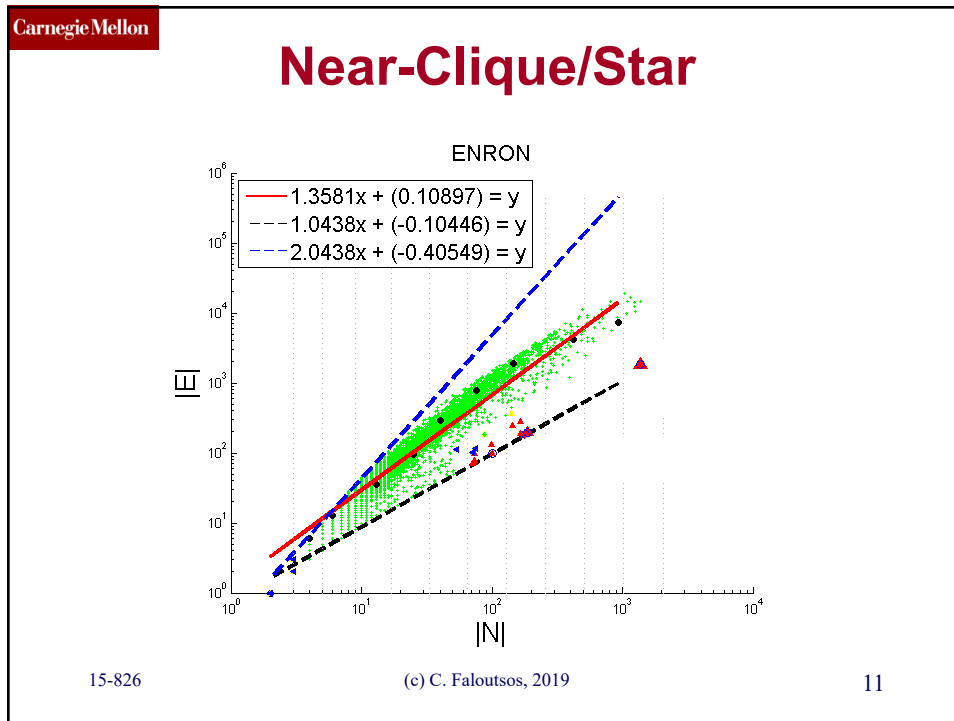
<http://instapundit.com/archives/025235.php>

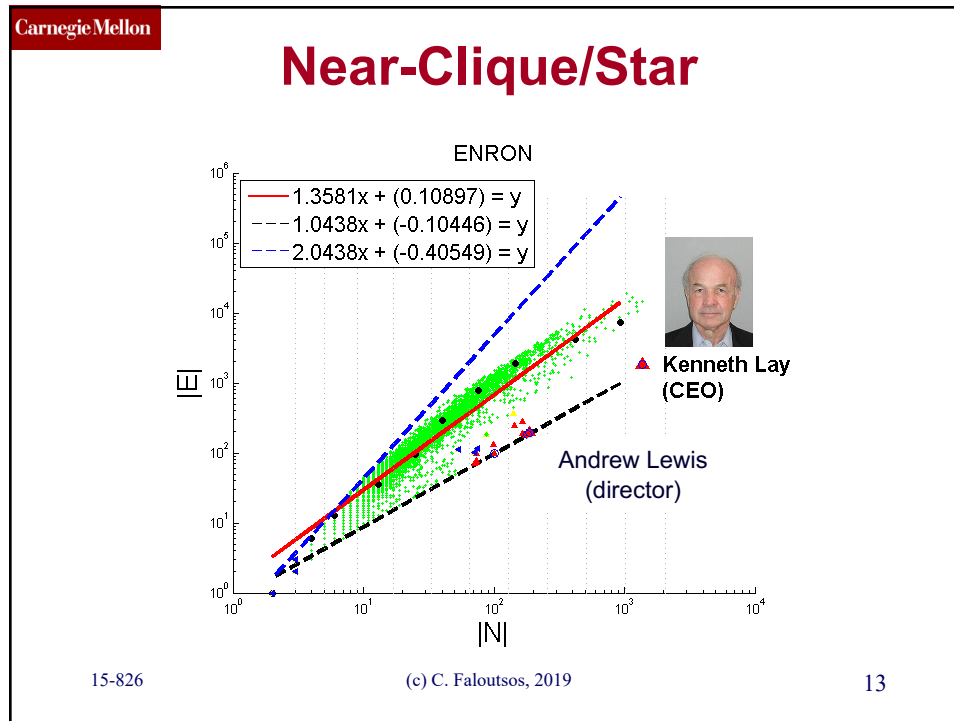
— $1.1094x + (-0.21414) = y$

- - - $1.1054x + (-0.21432) = y$

- - - $2.1054x + (-0.51535) = y$

15-826 10
(c) C. Faloutsos, 2019






CarnegieMellon

Roadmap

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

➔



15-826 (c) C. Faloutsos, 2019 14

CarnegieMellon

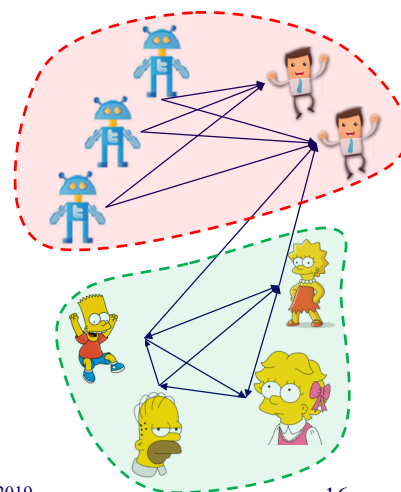
Catchsync: catch synchronized behavior in large directed graphs

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



CarnegieMellon

Fraud Detection: Graph Analysis Problem




15-826

(c) C. Faloutsos, 2019

16


CarnegieMellon

Fraud Detection: Graph Analysis Problem



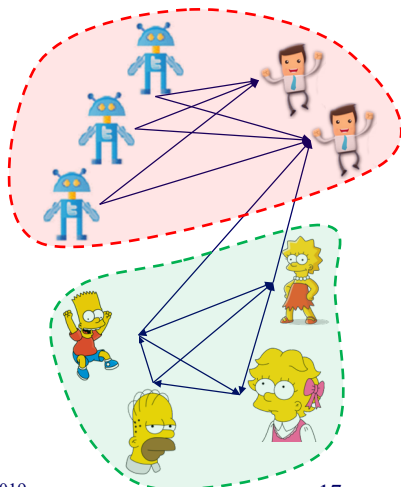
[www.buyfollowz.org]

5,000 FOLLOWERS \$69.99 Buy Now	2,000 FOLLOWERS \$29.99 Buy Now	1,000 FOLLOWERS \$15.99 Buy Now	10,000 FOLLOWERS \$119.99 Buy Now	20,000 FOLLOWERS \$229.99 Buy Now
---------------------------------------	---------------------------------------	---------------------------------------	---	---



[buymorelikes.com]


25,000 Facebook Likes \$265	50,000 Facebook Likes \$525	100,000 Facebook Likes \$1,000	200,000 Facebook Likes \$1,750
--------------------------------	--------------------------------	-----------------------------------	-----------------------------------



15-826
(c) C. Faloutsos, 2019
17


CarnegieMellon

Fraud Detection: Graph Analysis Problem



[buycheaplikes.com]

PACK-1 YOUTUBE 150 Real YouTube Likes \$5.00 (USD)	PACK-2 YOUTUBE 300 Real YouTube Likes \$9.00 (USD)	PACK-3 YOUTUBE 500 Real YouTube Likes \$13.00 (USD)	PACK-4 YOUTUBE 1,000 Real YouTube Likes \$25.00 (USD)
---	---	--	--



[reviewsteria.com]

It's easy to buy Amazon reviews. Just choose the number of reviews you would like to receive.

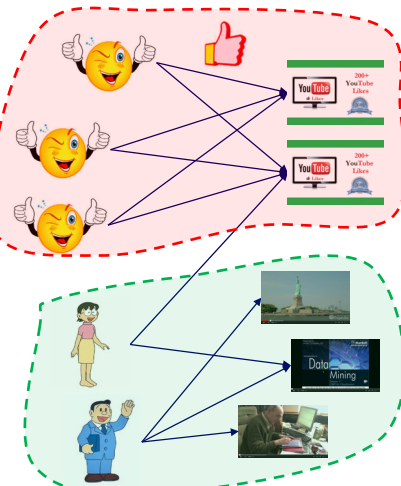
High quality reviews that customers love. 100% unique content by native speaking professional writers.

Choose the number of reviews and click Buy Now button to ramp up your Amazon business NOW.

Choose the number of reviews:

20


Buy Now



15-826
(c) C. Faloutsos, 2019
18

CarnegieMellon

Behavior-based Features





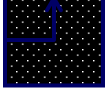
→ → → →

Follower behavior


Followee behavior


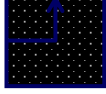
→ → → →







≈



≈



Out-degree
1st left singular vector
(Hubness)

...

In-degree
1st right singular vector
(Authoritativeness)

...

15-826
(c) C. Faloutsos, 2019
19

CarnegieMellon

Hubs/Authorities

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

(Later, in huge detail...)

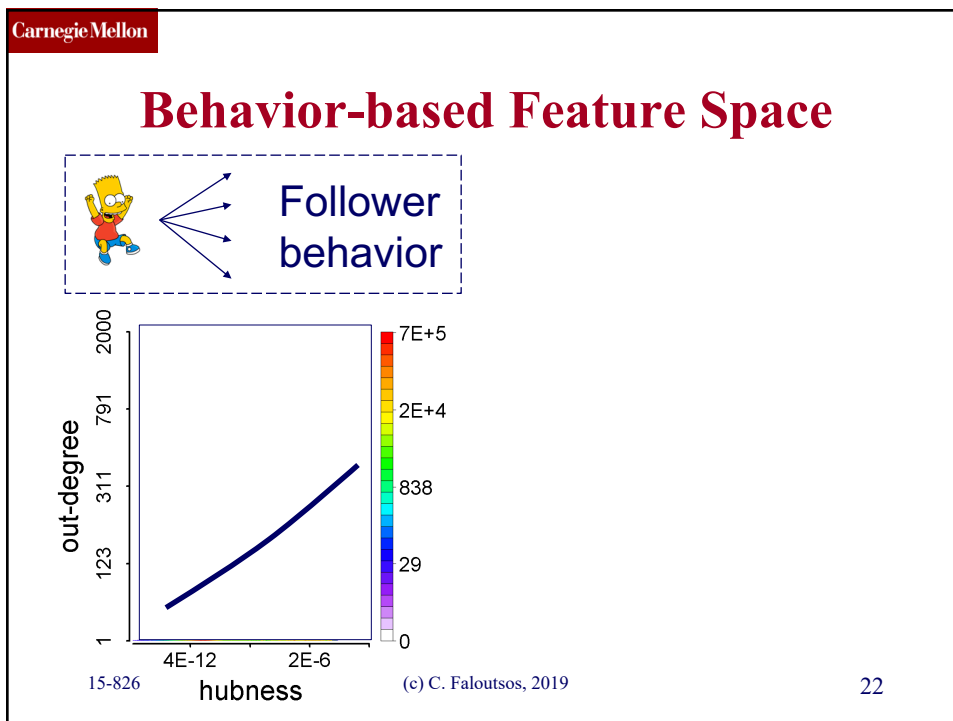
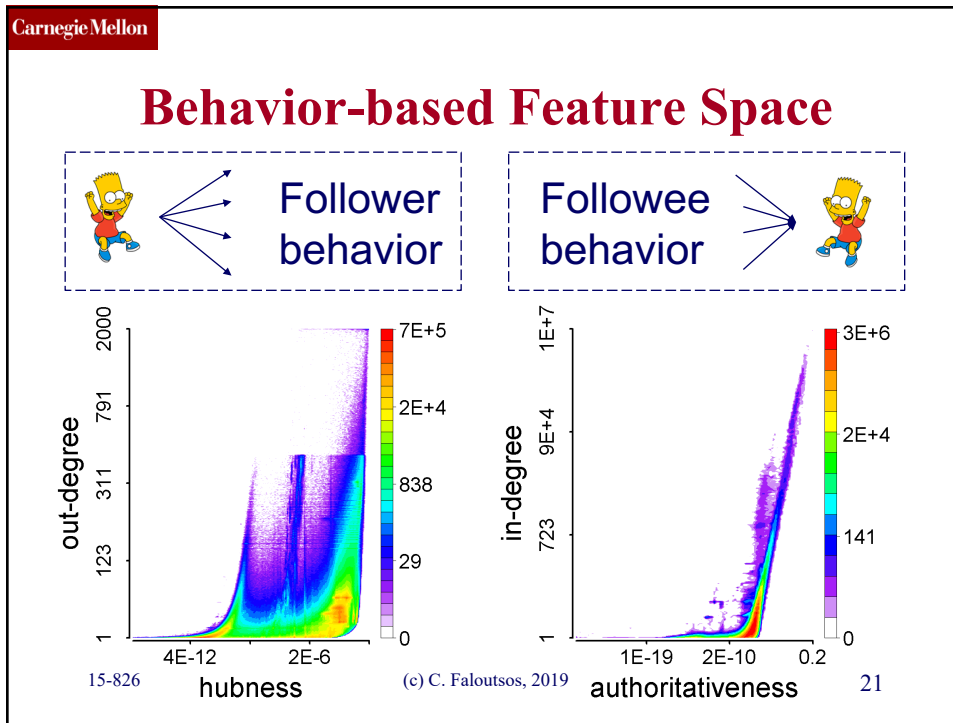
```

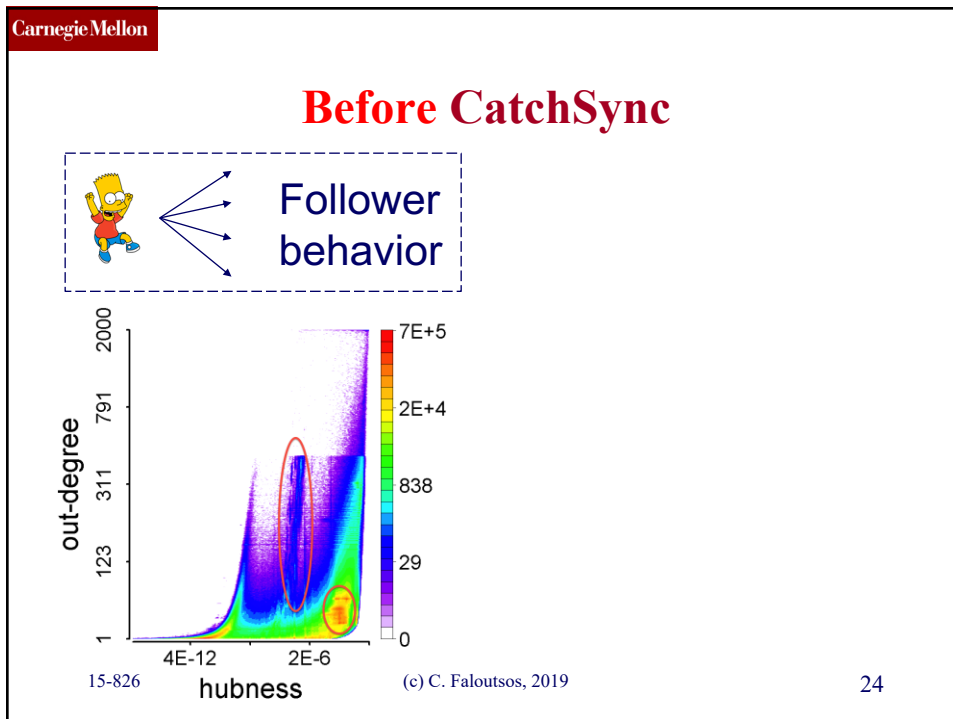
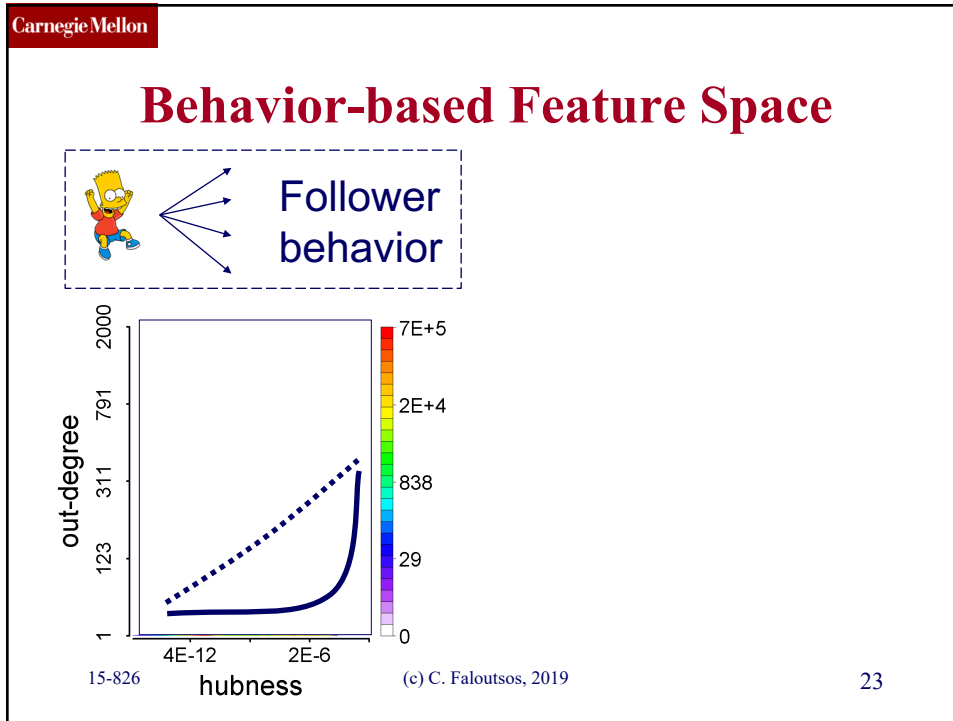
import networkx as nx
G=nx.path_graph(4)
h,a=nx.hits(G)

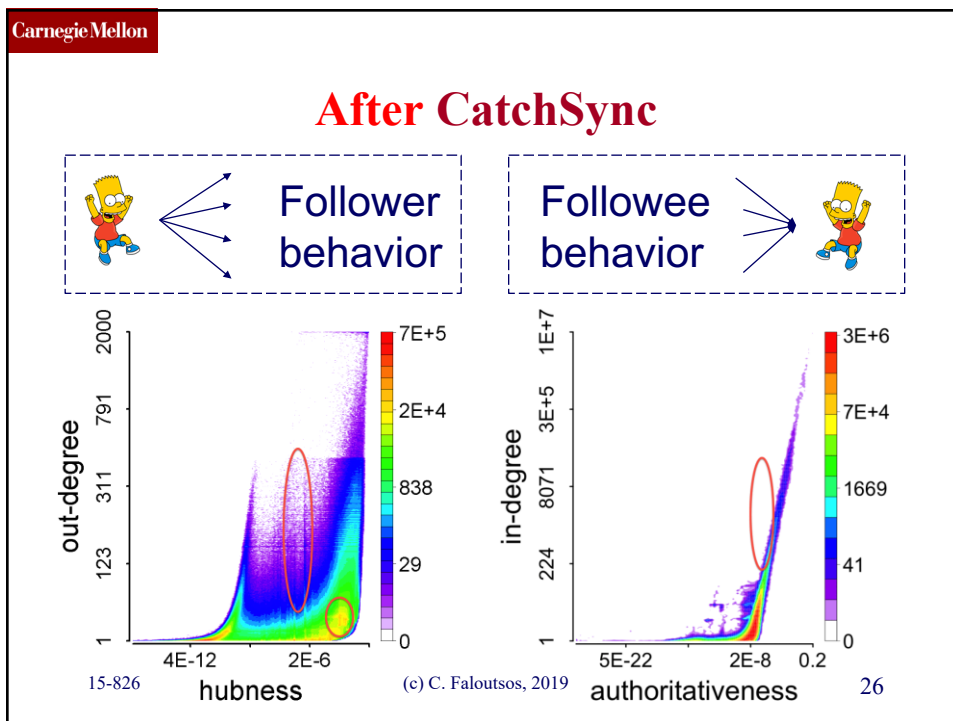
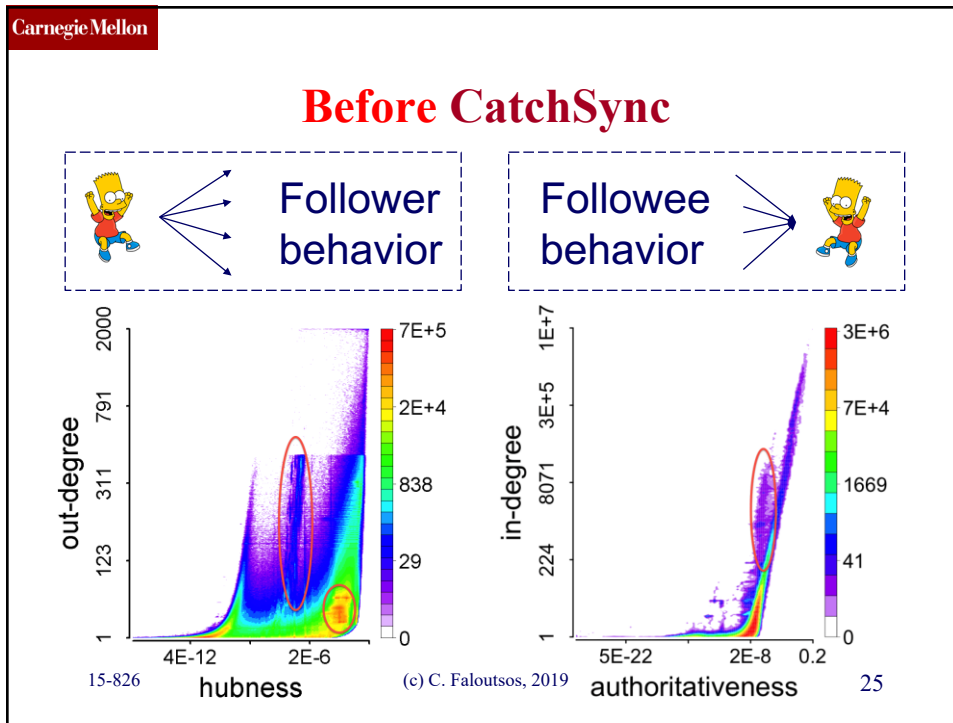
print h
print a

```

15-826
(c) C. Faloutsos, 2019
20









Conclusions

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

