



CATCHSYNC: CATCHING SYNCHRONIZED BEHAVIOR IN LARGE DIRECTED GRAPHS

Meng Jiang, Tsinghua University, Beijing, China

Joint work with Peng Cui, Alex Beutel,
Christos Faloutsos and Shiqiang Yang
August 26, 2014 – NYC, USA



Fraud Detection: Graph Analysis Problem



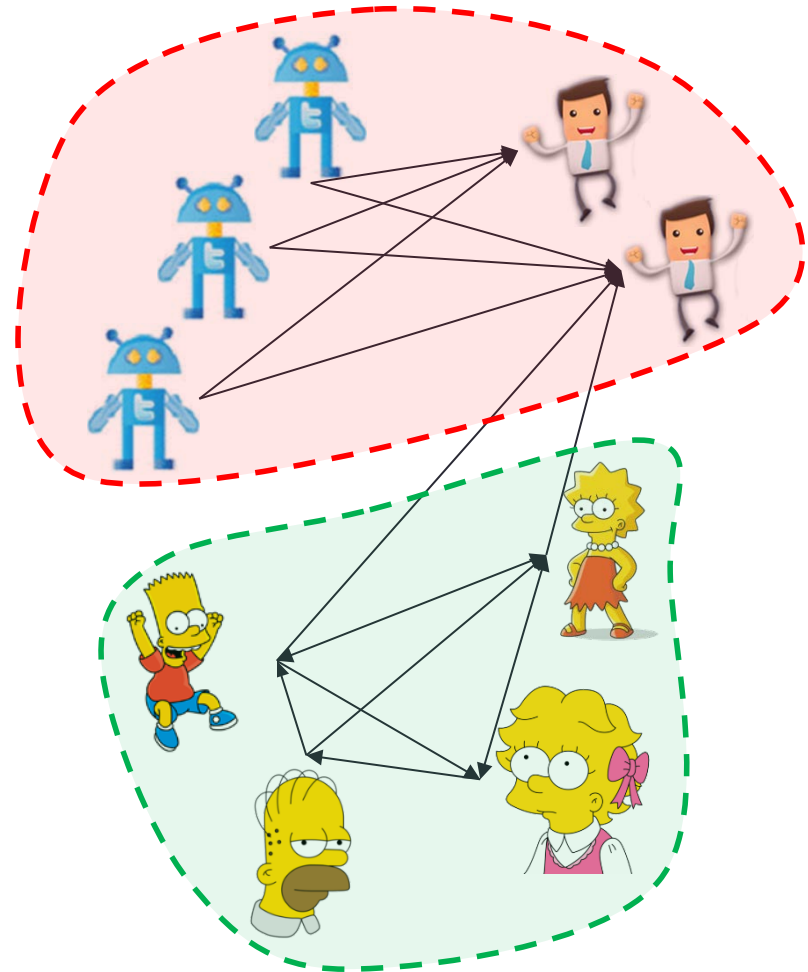
[www.buyfollowz.org]

5,000 FOLLOWERS \$69.99 Delivery within 3-4 days Buy Now VISA Save + 3%	2,000 FOLLOWERS \$29.99 Delivery within 2-3 days Buy Now VISA Save + 2%	1,000 FOLLOWERS \$15.99 Delivery within 1-2 days Buy Now VISA	10,000 FOLLOWERS \$119.99 Delivery within 4-5 days Buy Now VISA Save + 14%	20,000 FOLLOWERS \$229.99 Delivery within 5-8 days Buy Now VISA Save + 34%
--	--	---	---	---



[buymorelikes.com]

25,000 Facebook Likes \$265 Lifetime Replacement Warranty Dedicated 24/7 Customer Service 100% Risk Free, Try Us Today Order starts within 24 - 48 hours Order completed within 22 days	50,000 Facebook Likes \$525 Lifetime Replacement Warranty Dedicated 24/7 Customer Service 100% Risk Free, Try Us Today Order starts within 24 - 48 hours Order completed within 35 days	100,000 Facebook Likes \$1,000 Lifetime Replacement Warranty Dedicated 24/7 Customer Service 100% Risk Free, Try Us Today Order starts within 24 - 48 hours Order completed within 35 days	200,000 Facebook Likes \$1,750 Lifetime Replacement Warranty Dedicated 24/7 Customer Service 100% Risk Free, Try Us Today Order starts within 24 - 48 hours Order completed within 35 days
---	---	--	--



Fraud Detection: Graph Analysis Problem



[buycheaplikes.com]

PACK-1	PACK-2	PACK-3	PACK-4
YOUTUBE \$5	YOUTUBE \$9	YOUTUBE \$13	YOUTUBE \$25
150 Real YouTube Likes	300 Real YouTube Likes	500 Real YouTube Likes	1,000 Real YouTube Likes
\$ 5.00 (USD)	\$ 9.00 (USD)	\$ 13.00 (USD)	\$ 25.00 (USD)
Delivery within 24 hours Enter Your Video URL:	Delivery within 24-48 hours Please Enter Your Video URL:	Delivery within 24-48 hours Enter Your Video URL:	Delivery within 2-3 days Enter Your Video URL:
Current number of likes: <input type="text"/>	Current number of likes: <input type="text"/>	Current number of likes: <input type="text"/>	Current number of likes: <input type="text"/>
Add to Cart	Add to Cart	Add to Cart	Add to Cart



[reviewsteria.com]

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

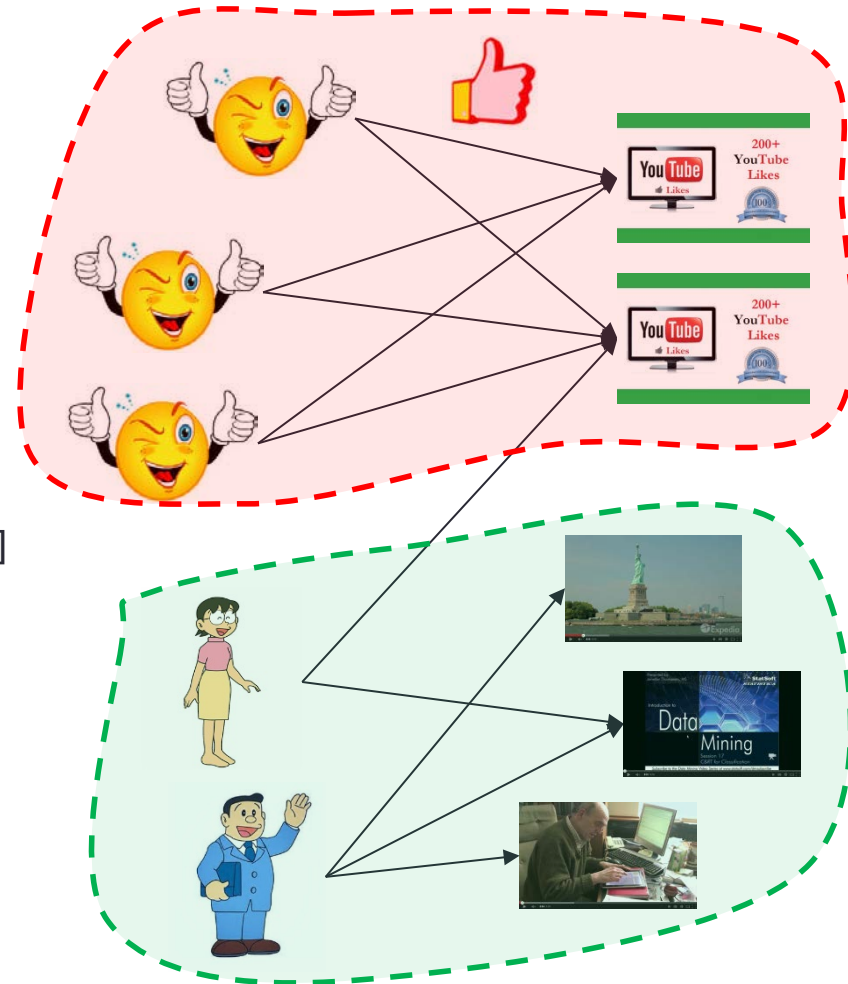
High quality reviews that customers love. 100% unique content by native speaking professional writers.

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](#)



Our Goals

- Given: A graph (large-scale, directed, etc.)
- Find: Frauds = Anomalous edges
- Goals:
 - G1. Find **patterns** that **distinguish** fraudsters from normal users
 - G2. Design **algorithms** that **catch** fraudsters

OUTLINE

1. Background

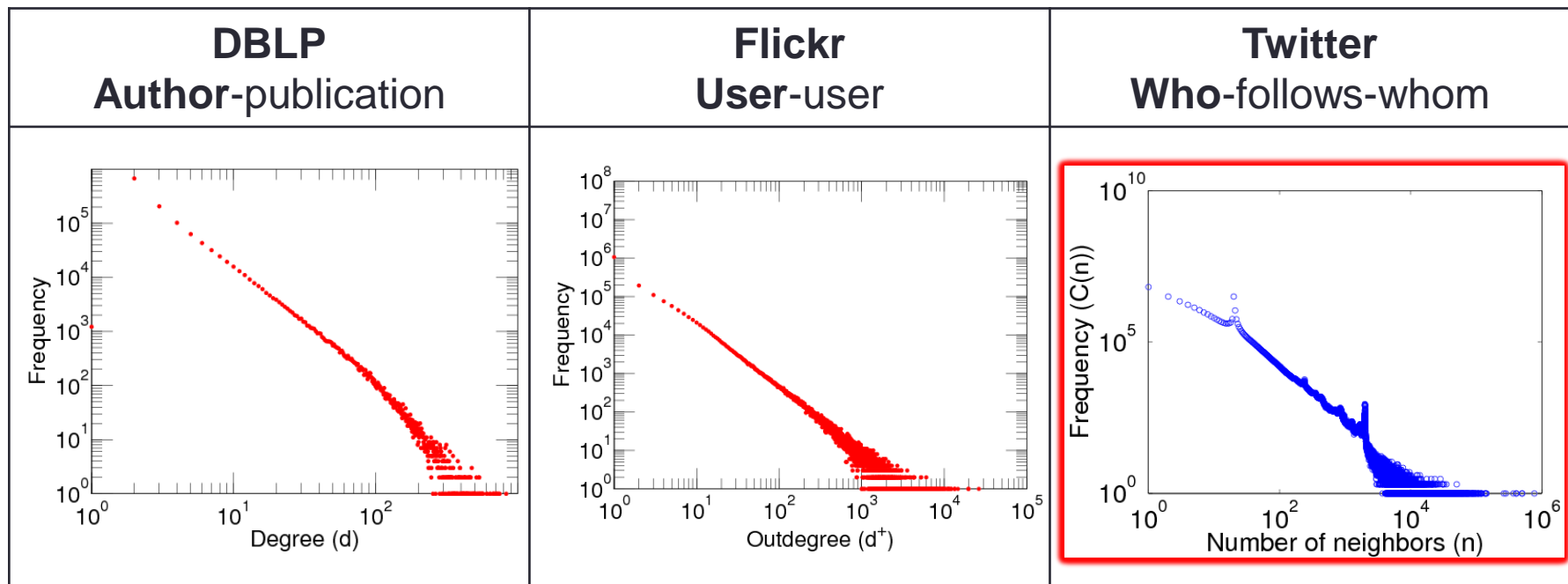
2. Fraudulent Pattern

3. The Algorithm

4. Experiments

Anomalies in Degree Distributions

- Power-law distribution



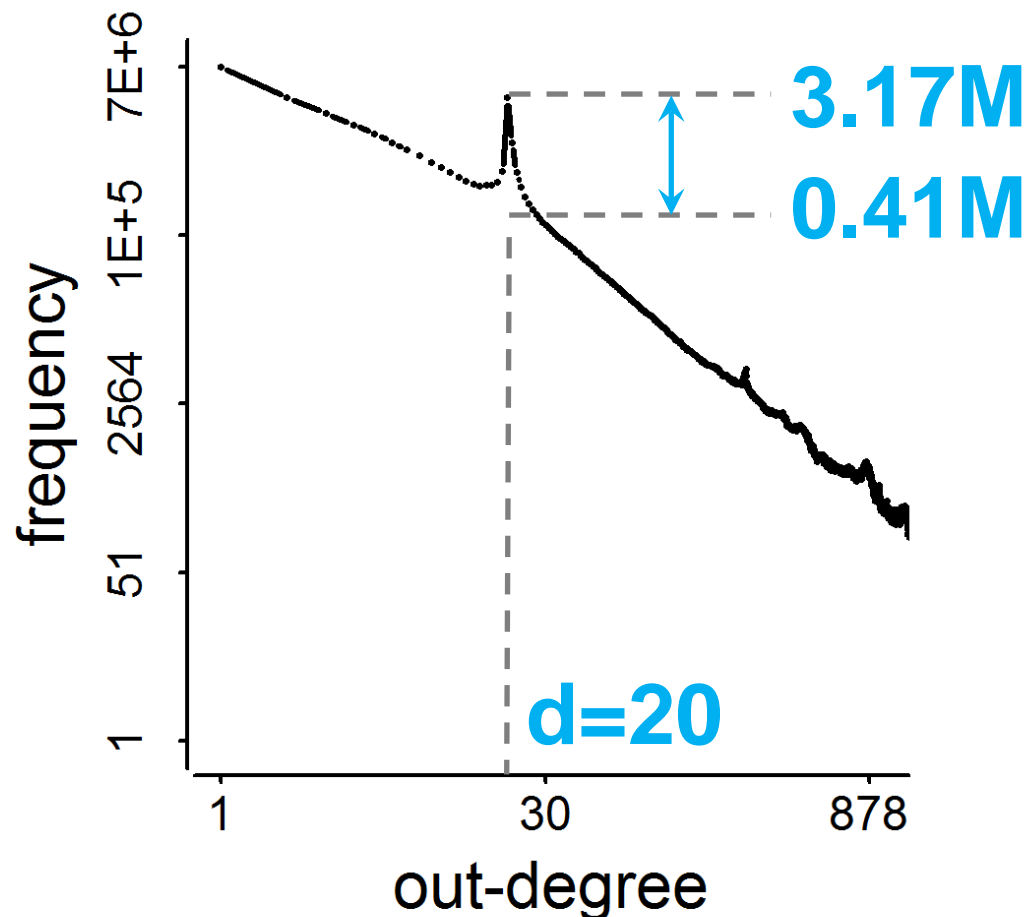
[konect.uni-koblenz.de/networks/]

Anomalies in Degree Distributions

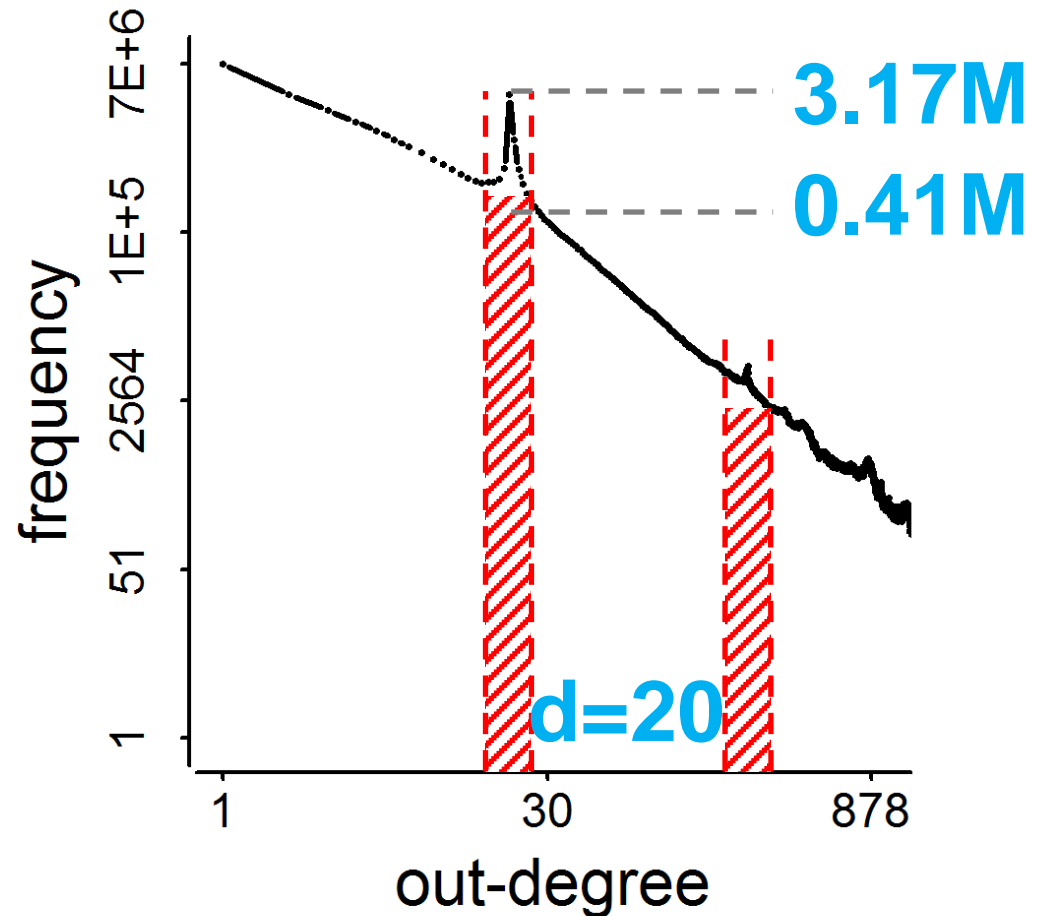
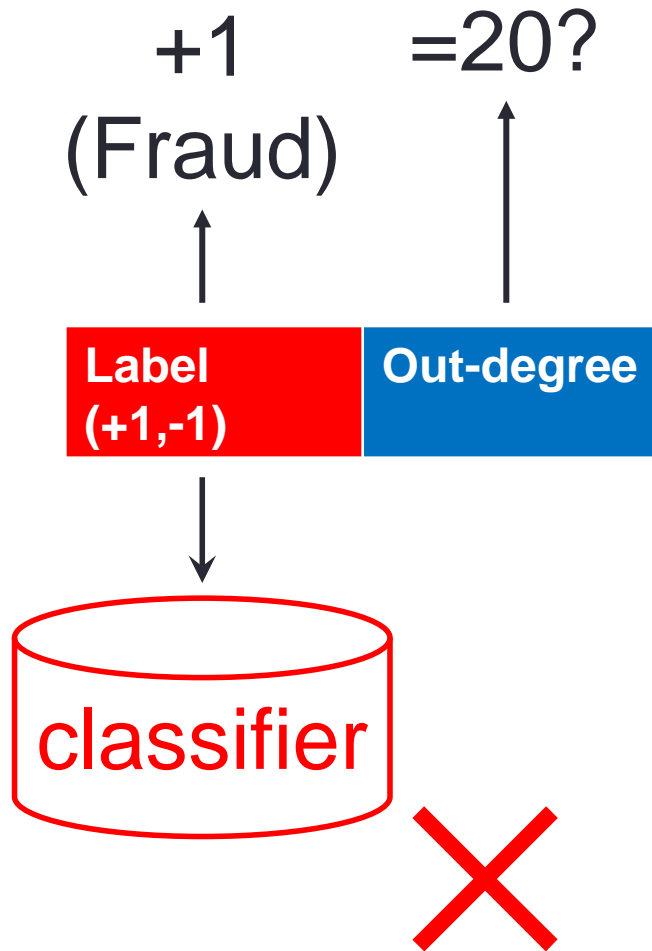


2009

41M



Linear Classifier with “Degree”: Fail

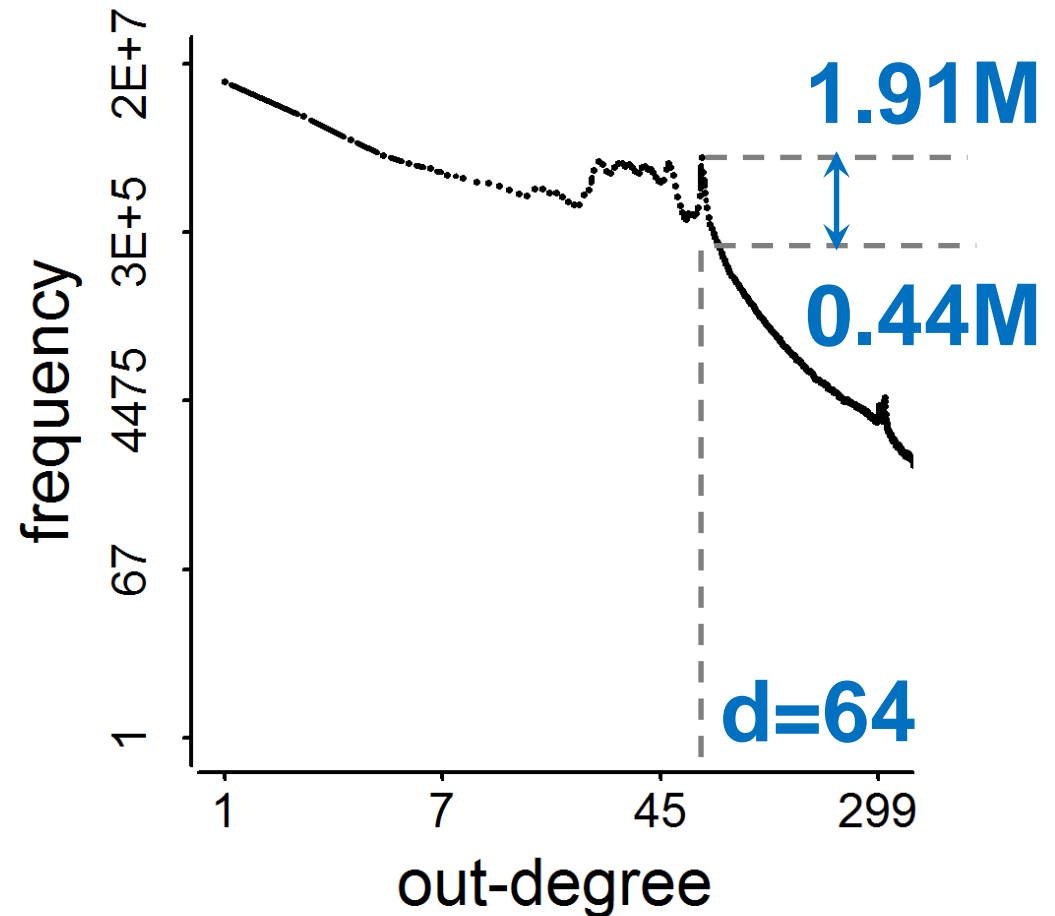


Graph Structure Distorted

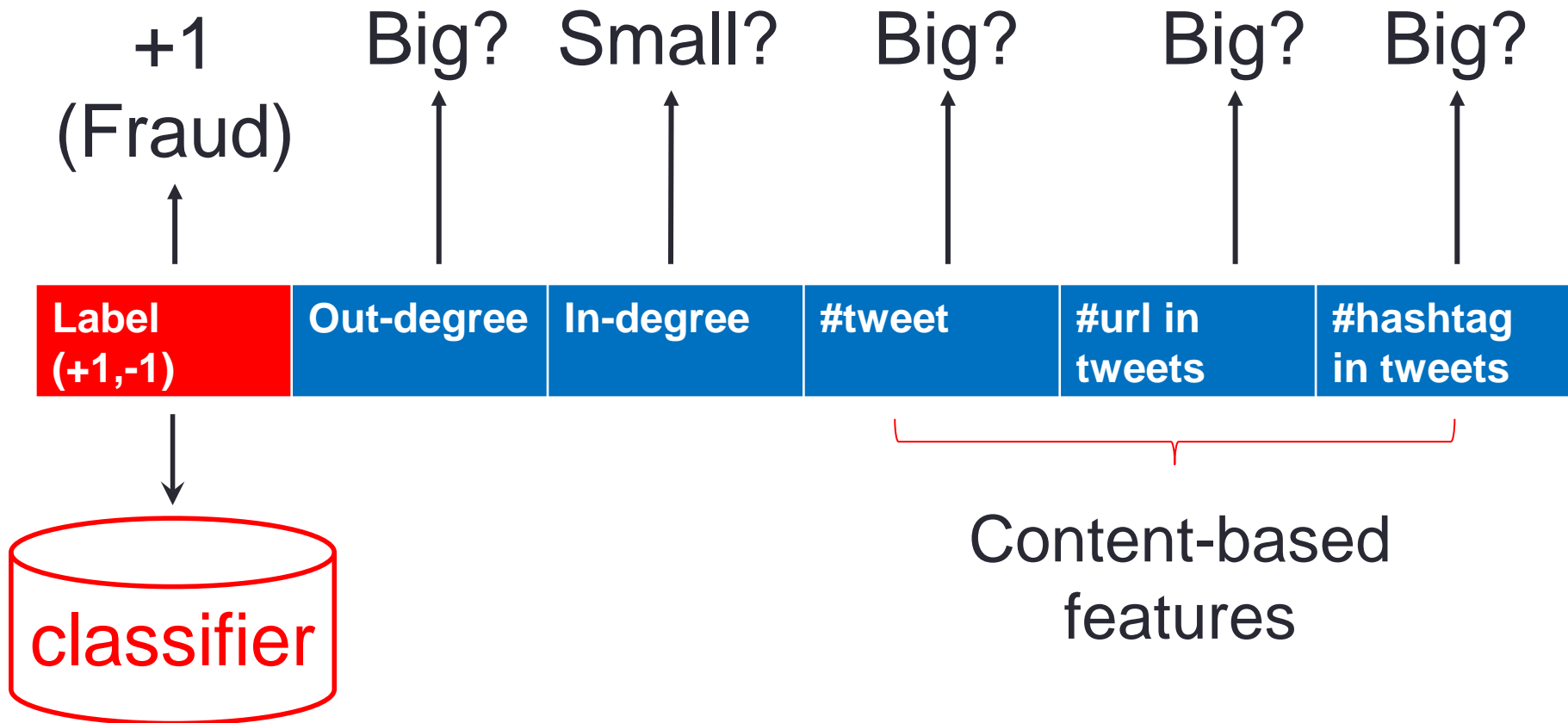


2011

117M



Traditional Fraud Detection




Empty Profile?





The image shows a screenshot of a Twitter profile for Annalisa Monsod. The profile header is dark with a red profile picture. The name 'Annalisa Monsod' is in white, followed by the handle '@AnnalisaMonsodz' and a 'FOLLOWS YOU' badge. Below this, it says 'By: Dan Schawbel on November 30th, 2010 at 11:19 am' and 'South Cambridgeshire, UK'. The stats bar shows 17 tweets, 2,000 following, and 12 followers. There is a 'Follow' button. The tweets section shows two tweets from July 6th, both with the same red profile picture. The first tweet is about an 'Entrepreneurial Spotlight' and the second is a correction about capitalization in a title.

Annalisa Monsod
@AnnalisaMonsodz FOLLOWS YOU
By: Dan Schawbel on November 30th, 2010 at 11:19 am
South Cambridgeshire, UK


17 TWEETS 2,000 FOLLOWING 12 FOLLOWERS 

Tweets

 **Annalisa Monsod** @AnnalisaMonsodz 6 Jul
Entrepreneurial Spotlight: Daniel Brusilovsky - Grasshopper Group
Expand


 **Annalisa Monsod** @AnnalisaMonsodz 6 Jul
The only place I see that he?s typed it as ?MAC? is in the titles, in which case all of the letters are capitalized, as it?s a title
Expand

Few Followers?




AjaQwX1Z3
@AjaQwX1Z3

0 TWEETS 741 FOLLOWING 344 FOLLOWERS




nikhil_parekh
@nikhil_parekh

Internet Marketer, Affiliate Marketer, World Class Public Speaker, Trainer, Coach & Mentor to Thousands




Anthony Gemma ✓
@AnthonyGemma

President, CEO, Candidate for US Congress, Radio and TV Talk Show Host, Social Media Guru, Author, Innovator, Creative Marketing...




IDrive Online Backup
@IDriveBackup

Online backup to protect your digital life! 5 GB free starter space. Sign up today at IDrive.com!




Danny Whitehouse
@dannywhitehouse



Mia Rose
@DrMiaRose










Psychologist // Prize-winning author of 'Awaken to Love' // Publisher of Soulwoman eMagazine // Founder of the Soulwoman Sanctuary.



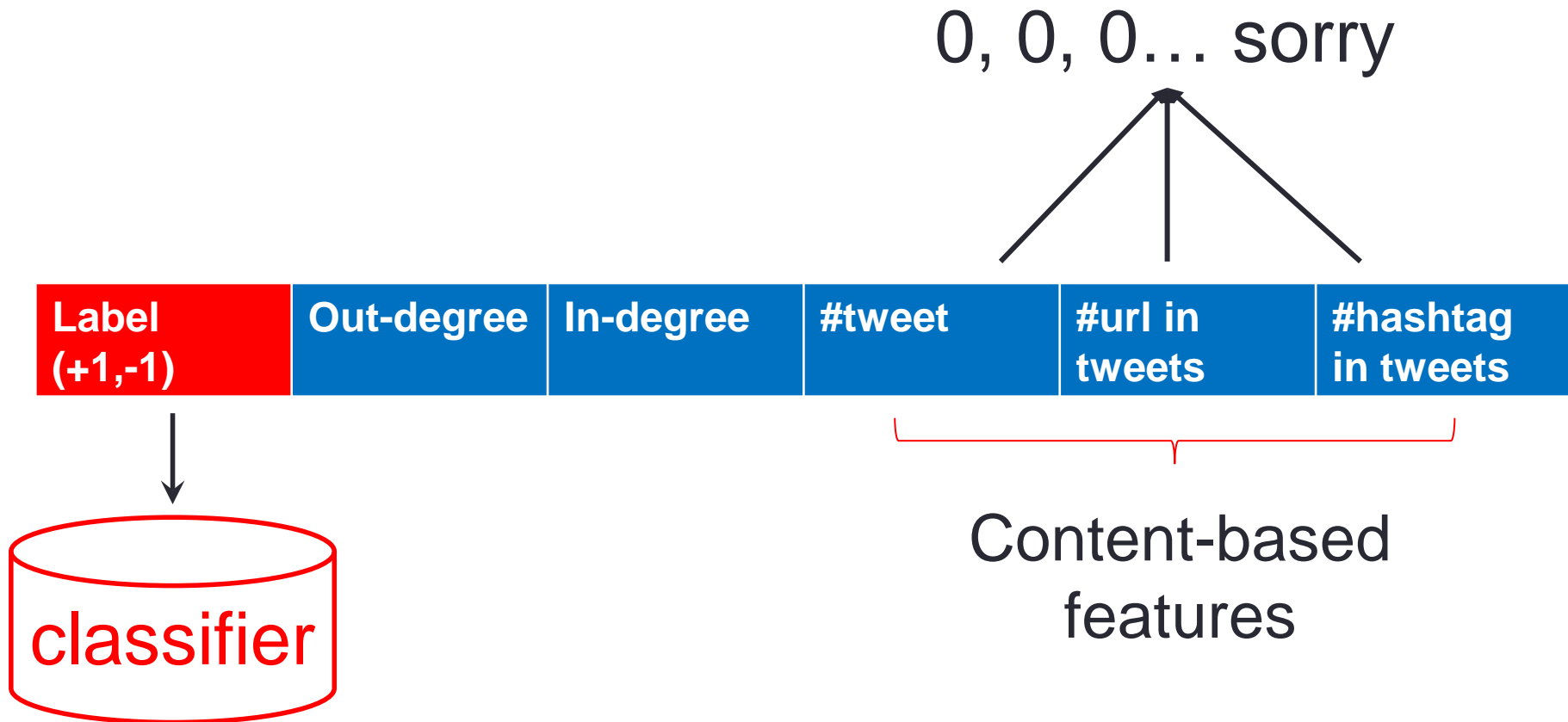
Kevin Cottrell
@kevincottrell

Austin REALTOR(R) with Prudential Texas Realty. Entrepreneur originally from Silicon Valley. Passionate about all things real estate and living...

Many Followings?

	Buy AB22 Propertwee @Buy_AB22	0 TWEETS	20 FOLLOWING	2 FOLLOWERS
 B.J. Mendelson @BJMendelson I just post silly stuff that I found on the Internet. None of it is mine.	 someecards @someecards Welcome to the Twitter feed of somewhat acclaimed humor sites, Someecards.com and HappyPlace.com . You'll love not unfollowing us!	 Steven Johnson @stevenbjohnson Author. (Latest: Future Perfect.) TV host. (How We Got To Now, on PBS/BBC soon.) Startup creator. (FEED, outside.in.) Dad. (Three boys.)...		
 adventuregirl @adventuregirl Hi Everyone! I'm Stef Michaels- an avid lifestyles journalist, TV personality, adventurer. Co-founder of KEEN Digital Summit. Contributor to Yahoo! Travel.	 People magazine @peplemag PEOPLE.com is the No. 1 site for celebrity news! Tweet your questions to our customer service team @Peoplemag_Help	 ashton kutcher @aplusk I make stuff, actually I make up stuff, stories mostly, collaborations of thoughts, dreams, and actions. Thats me.		
 Paul Pierce @paulpierce34 The one and only Truth. Founder of The @TruthonHealth.	 Mashable @mashable News, resources, inspiration and fun for the connected generation. Tweets by			

Content: Unavailable? Look Normal?



Behavior is the Key

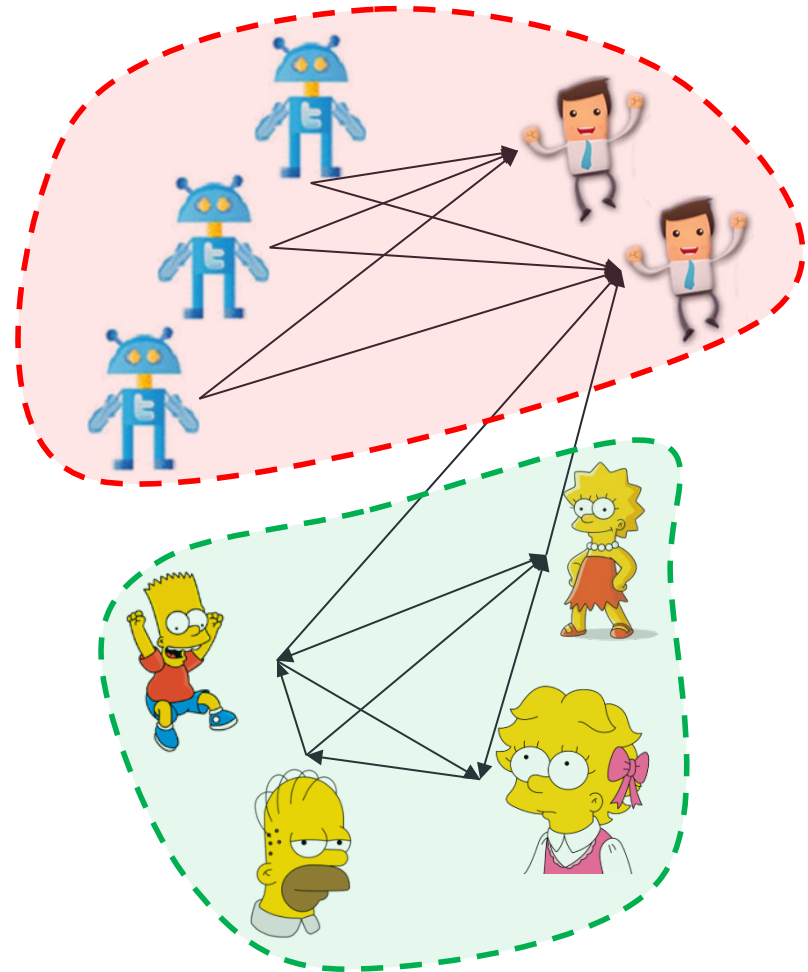
Monetary Incentive

Content

what they
appear to
behave

**Behavior/
Links**

what they
have to
behave



OUTLINE

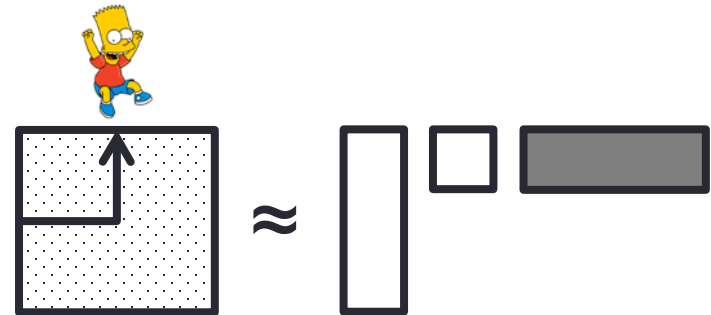
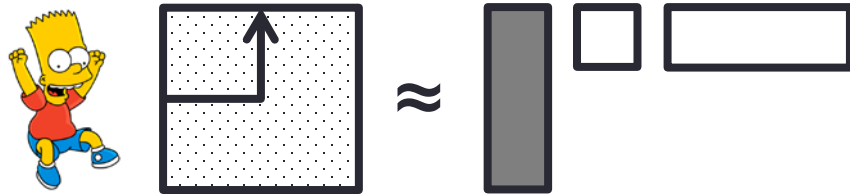
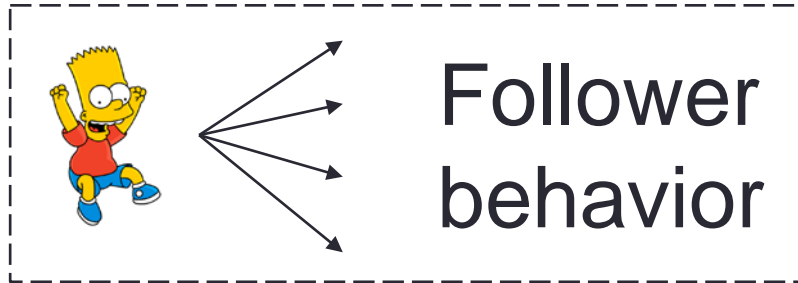
1. Background

2. Fraudulent Pattern

3. The Algorithm

4. Experiments

Behavior-based Features



Out-degree

1st left singular vector
(Hubness)

2nd left singular vector

...

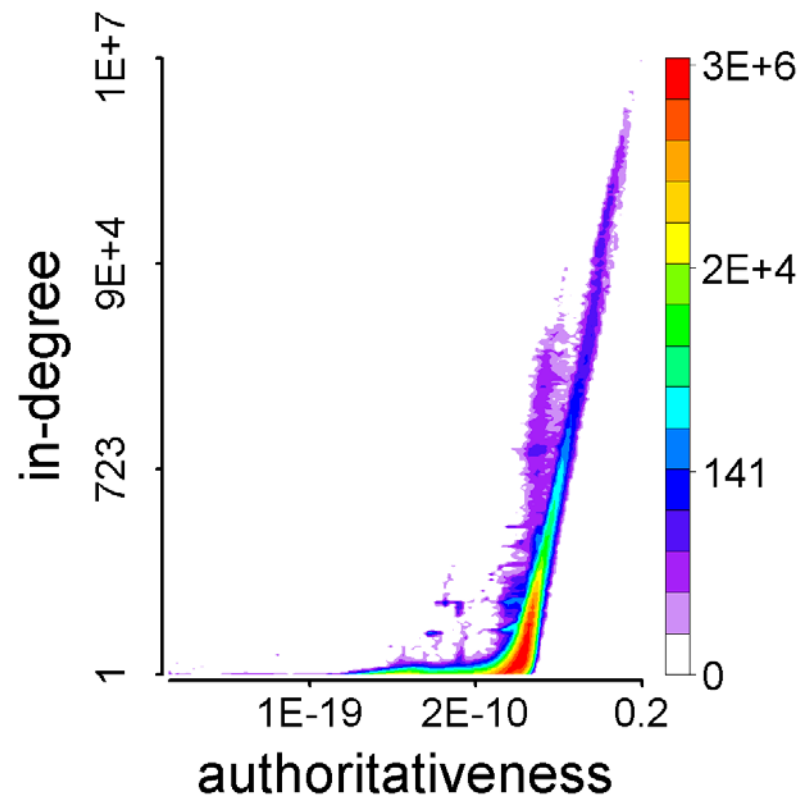
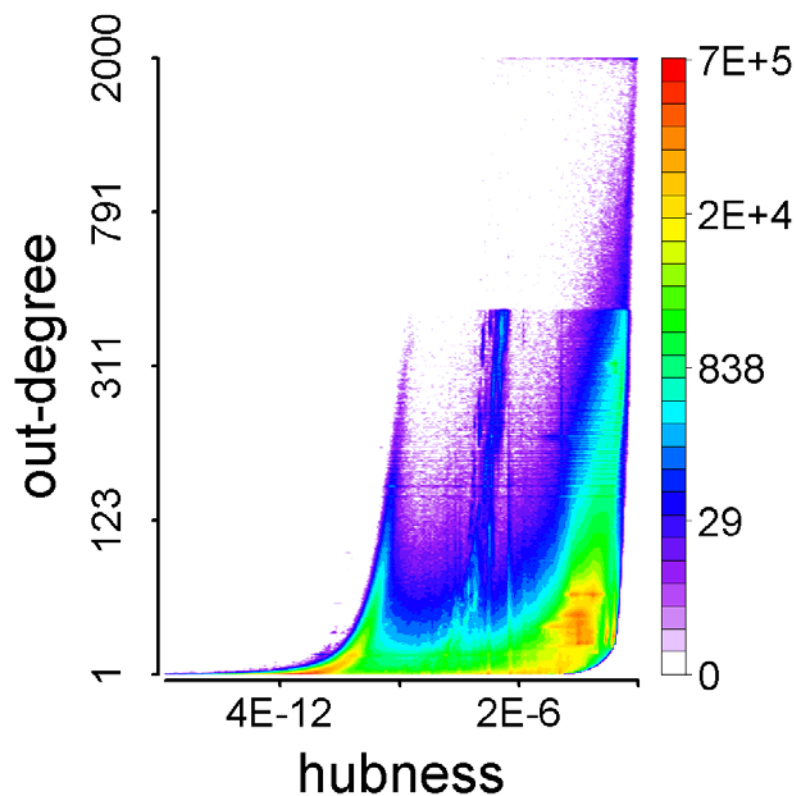
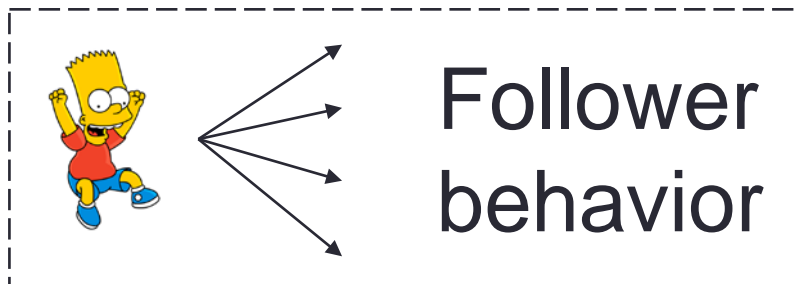
In-degree

1st right singular vector
(Authoritativeness)

2nd right singular vector

...

Behavior-based Feature Space



Fraudulent Behavior Patterns

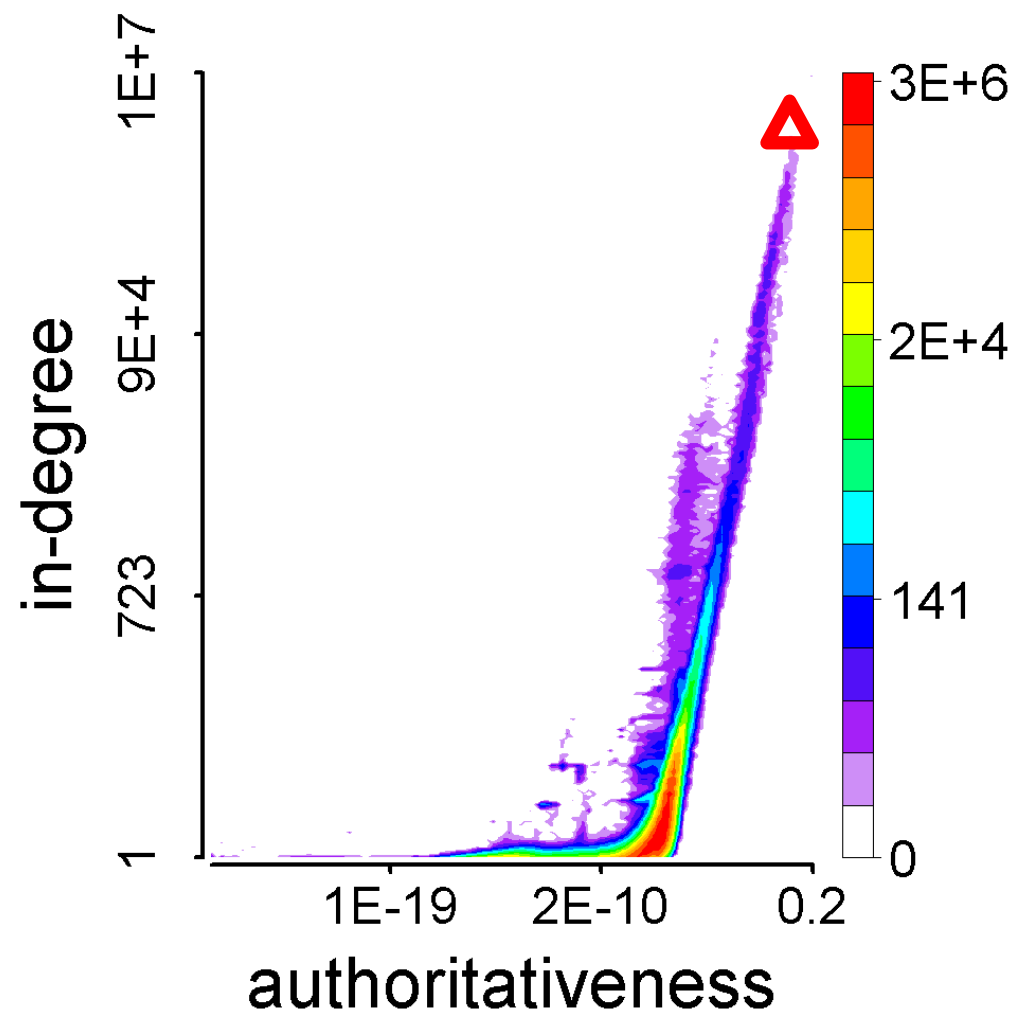


FOLLOWING
117

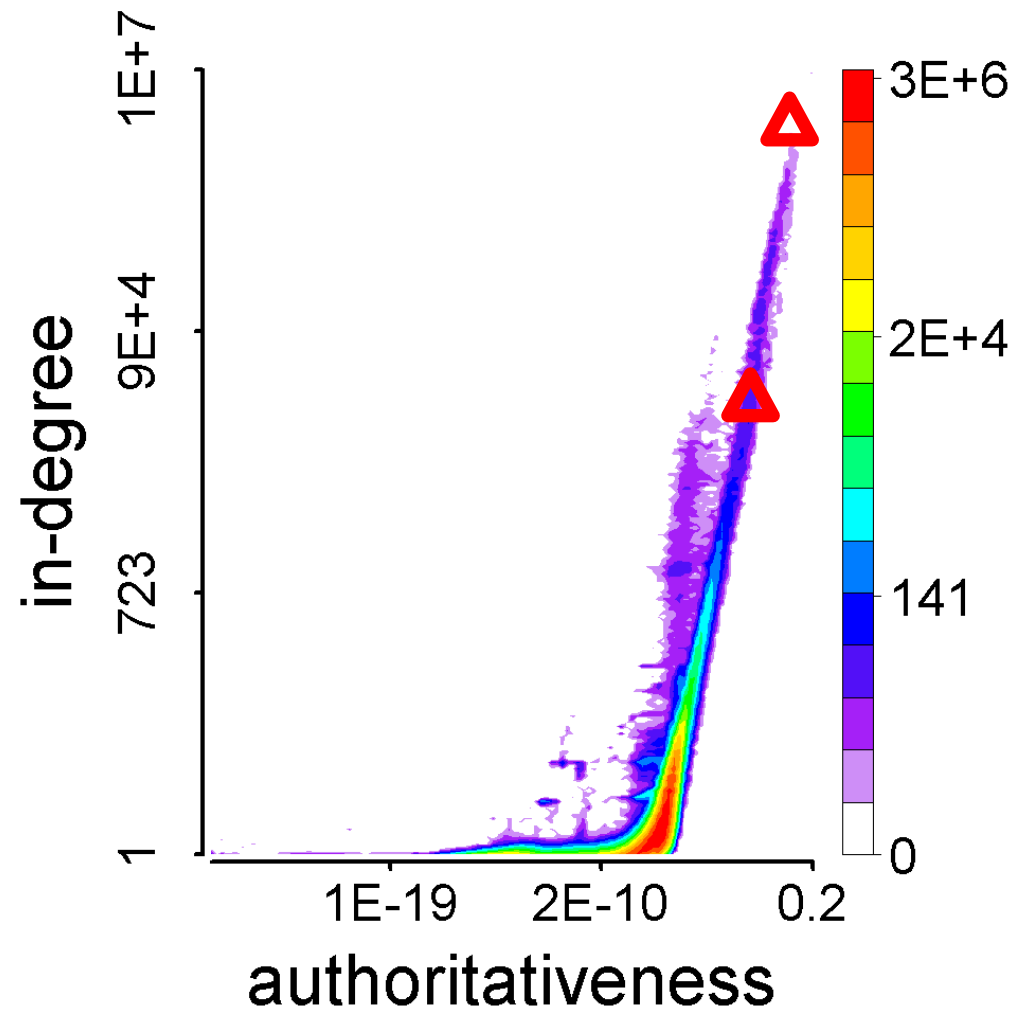


FOLLOWERS
16.8M

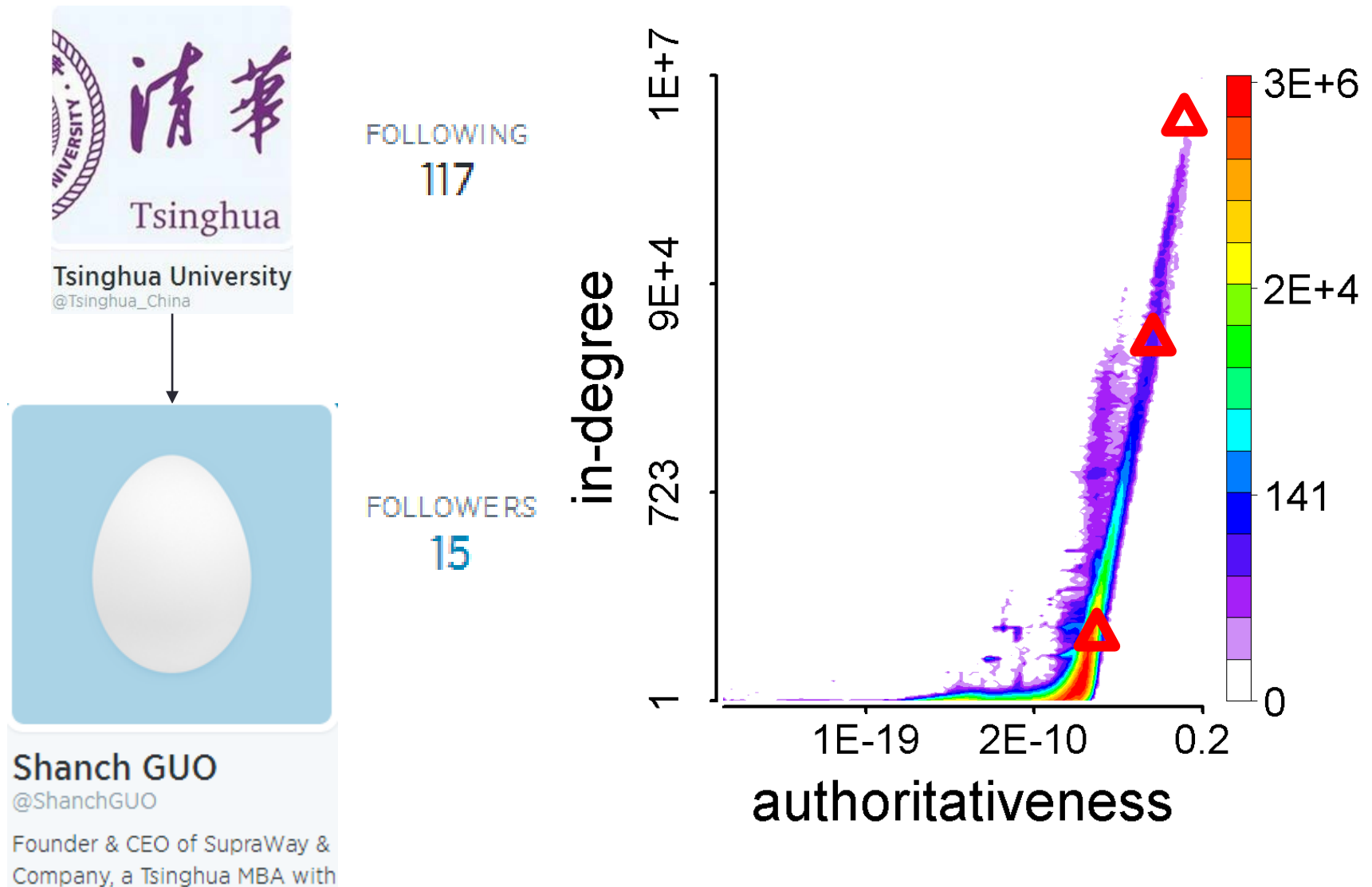
Bill Gates ✓
@BillGates



Fraudulent Behavior Patterns



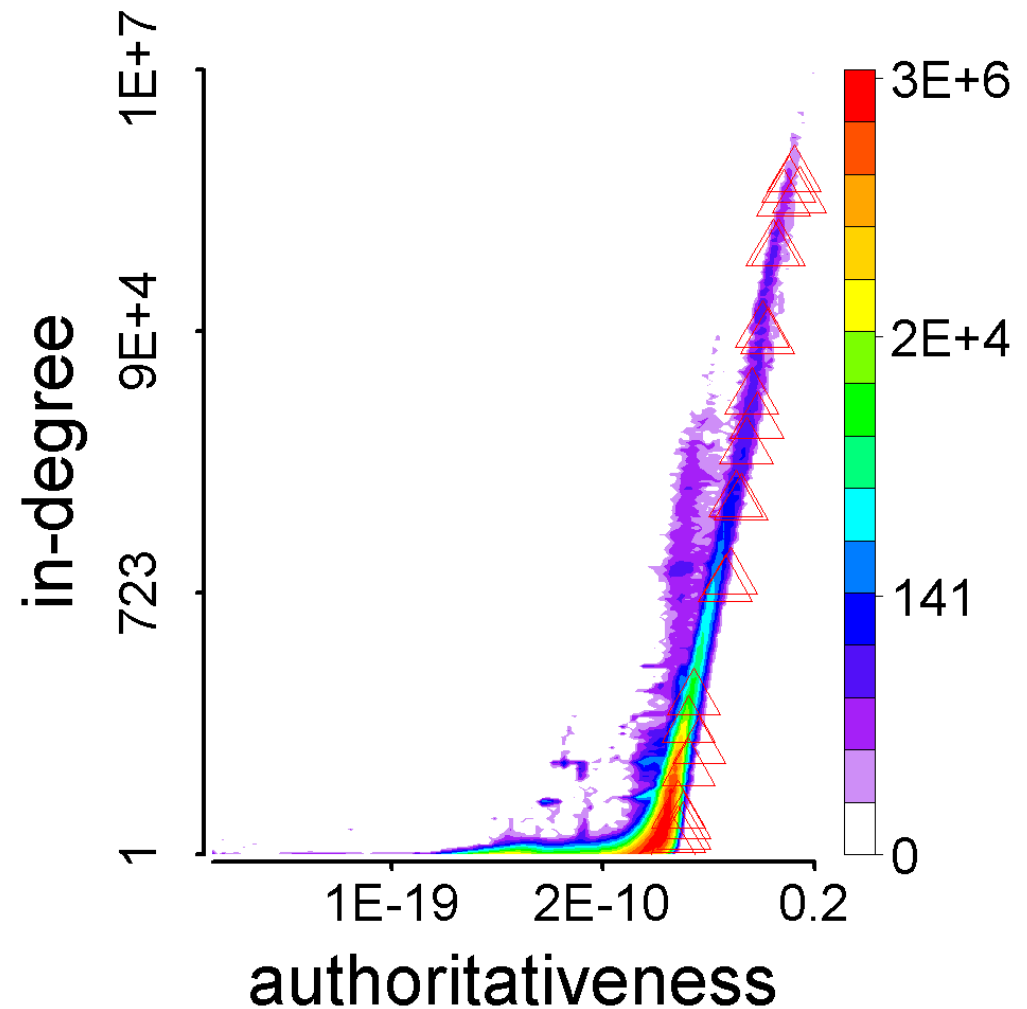
Fraudulent Behavior Patterns



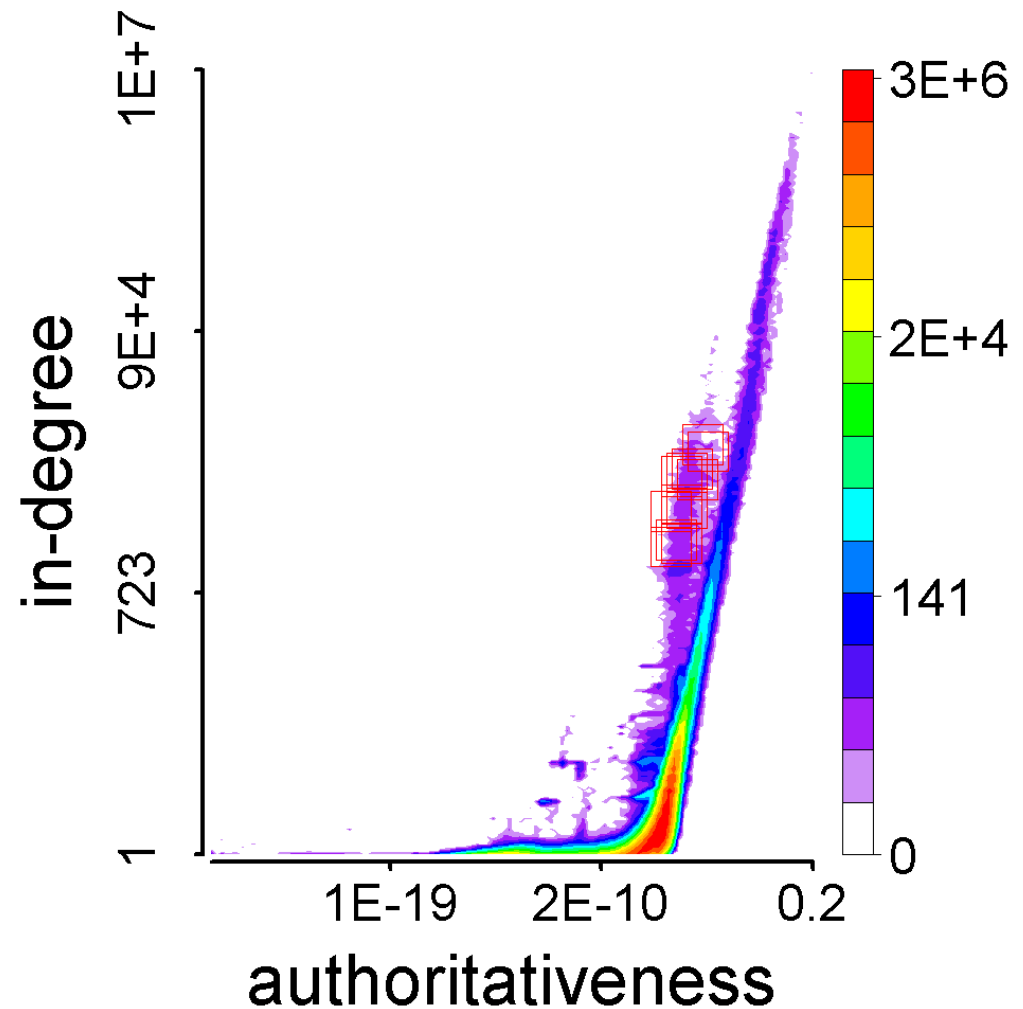
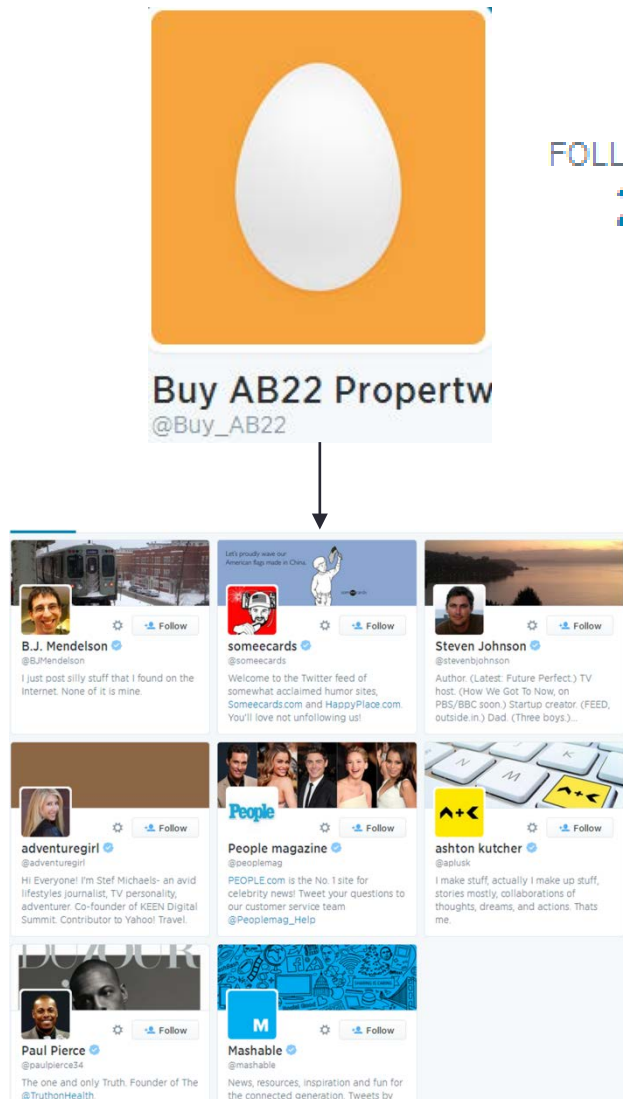
Fraudulent Behavior Patterns



FOLLOWING
117

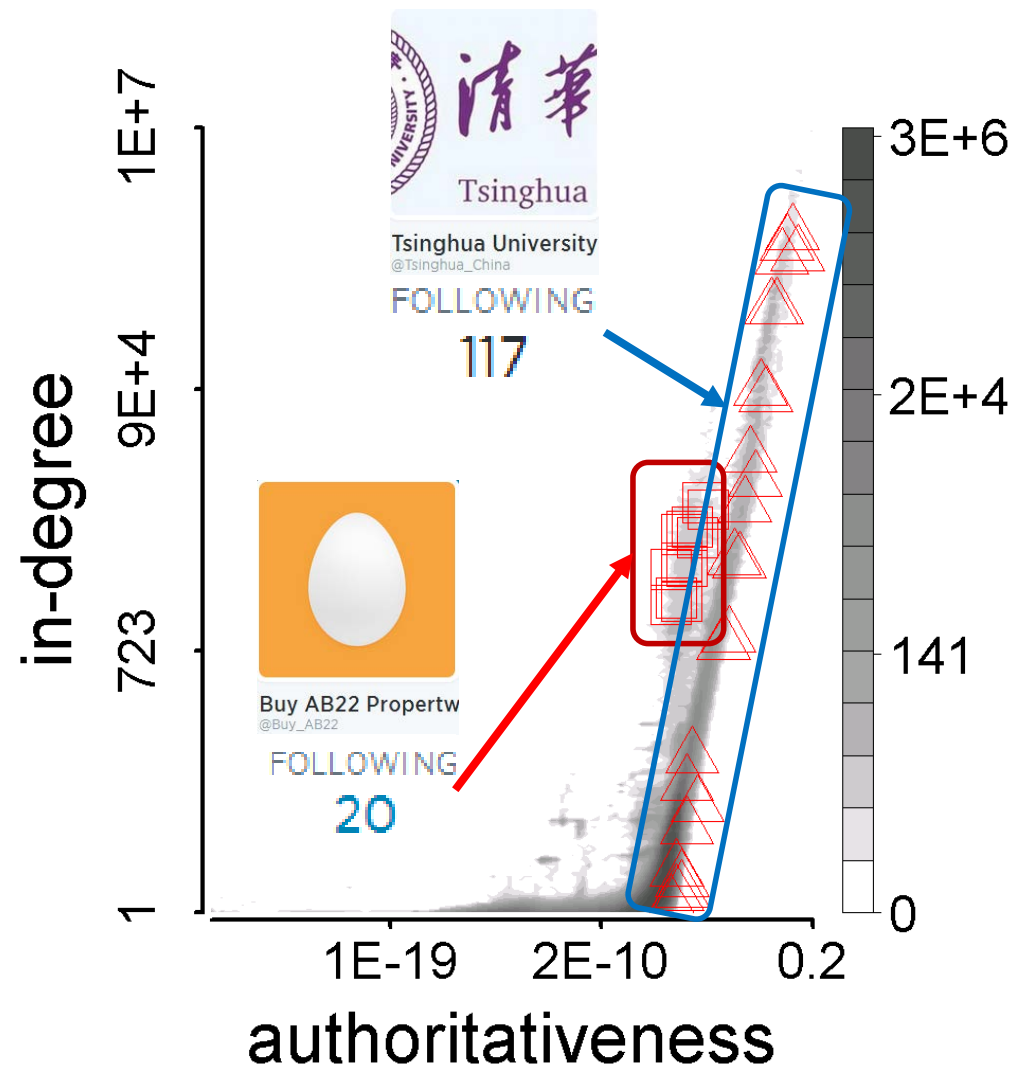


Fraudulent Behavior Patterns



Fraudulent Behavior Patterns

- Synchronized
- Abnormal



OUTLINE

1. Background

2. Fraudulent Pattern

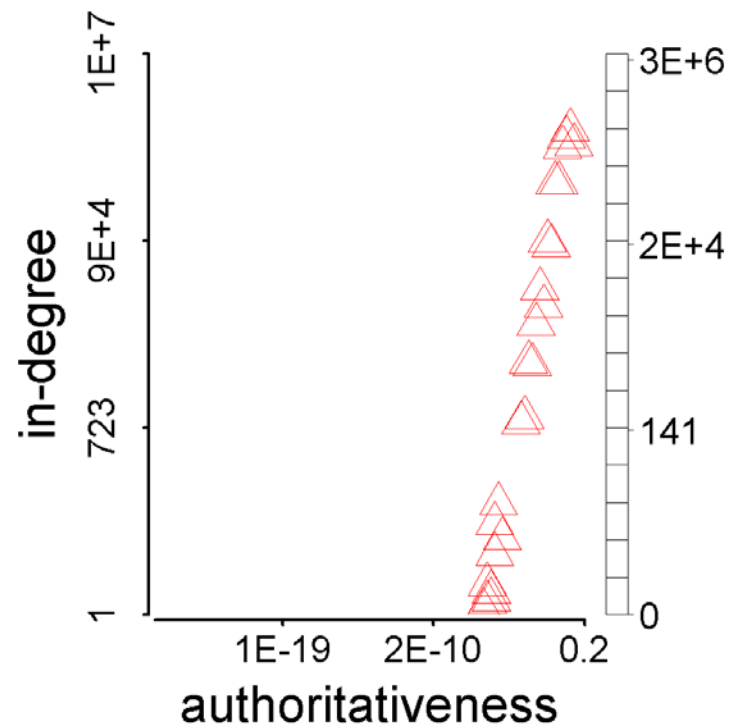
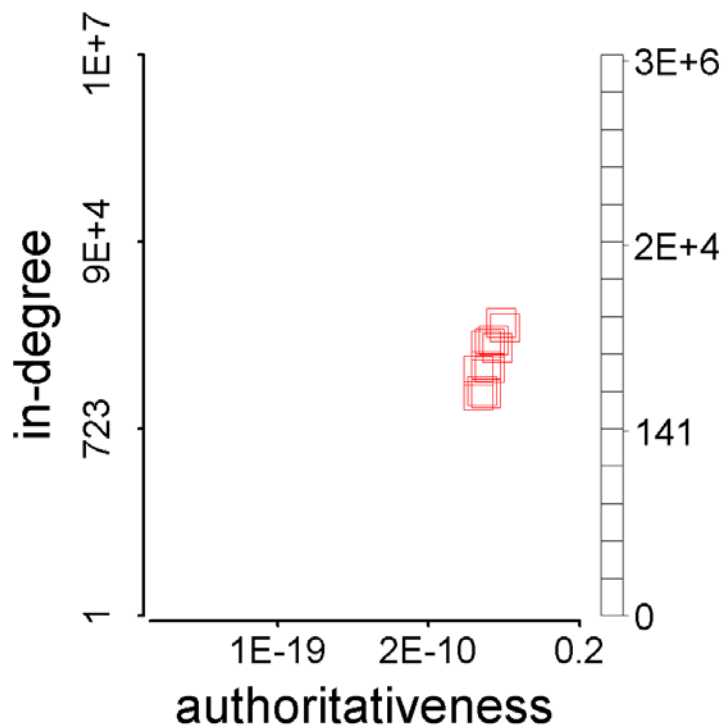
3. The Algorithm

4. Experiments

Synchronicity and Normality

- Synchronicity

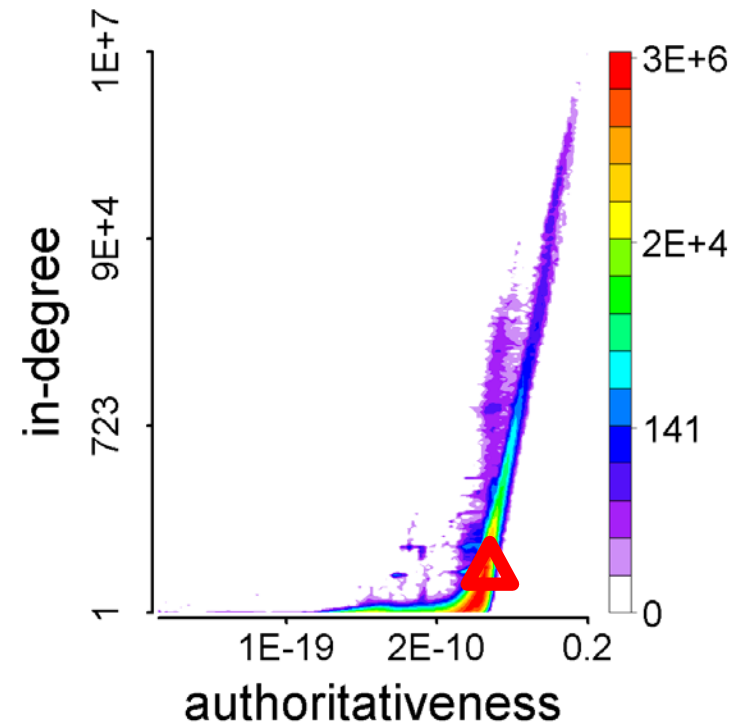
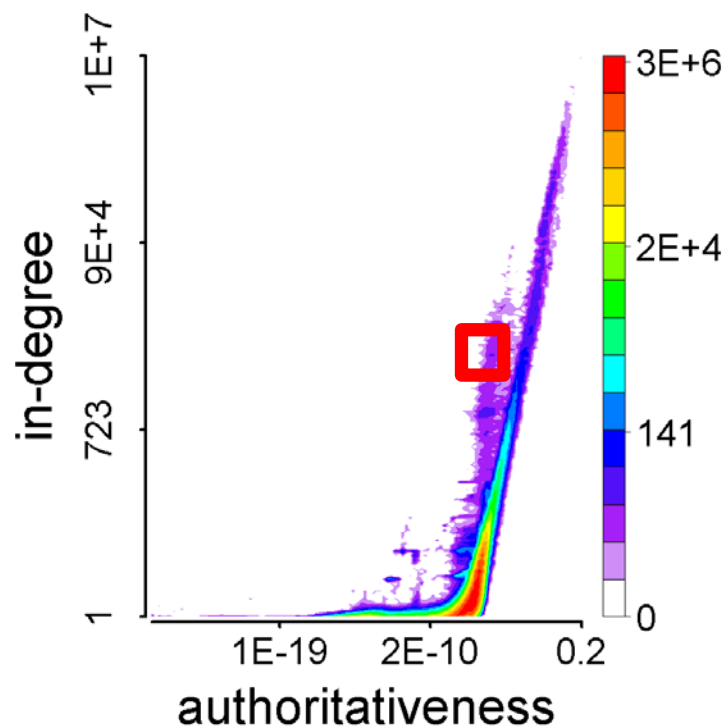
$$\text{sync}(u) = \frac{\sum_{(v,v') \in \mathcal{F}(u) \times \mathcal{F}(u)} \mathbf{p}(v) \cdot \mathbf{p}(v')}{d(u) \times d(u)}$$



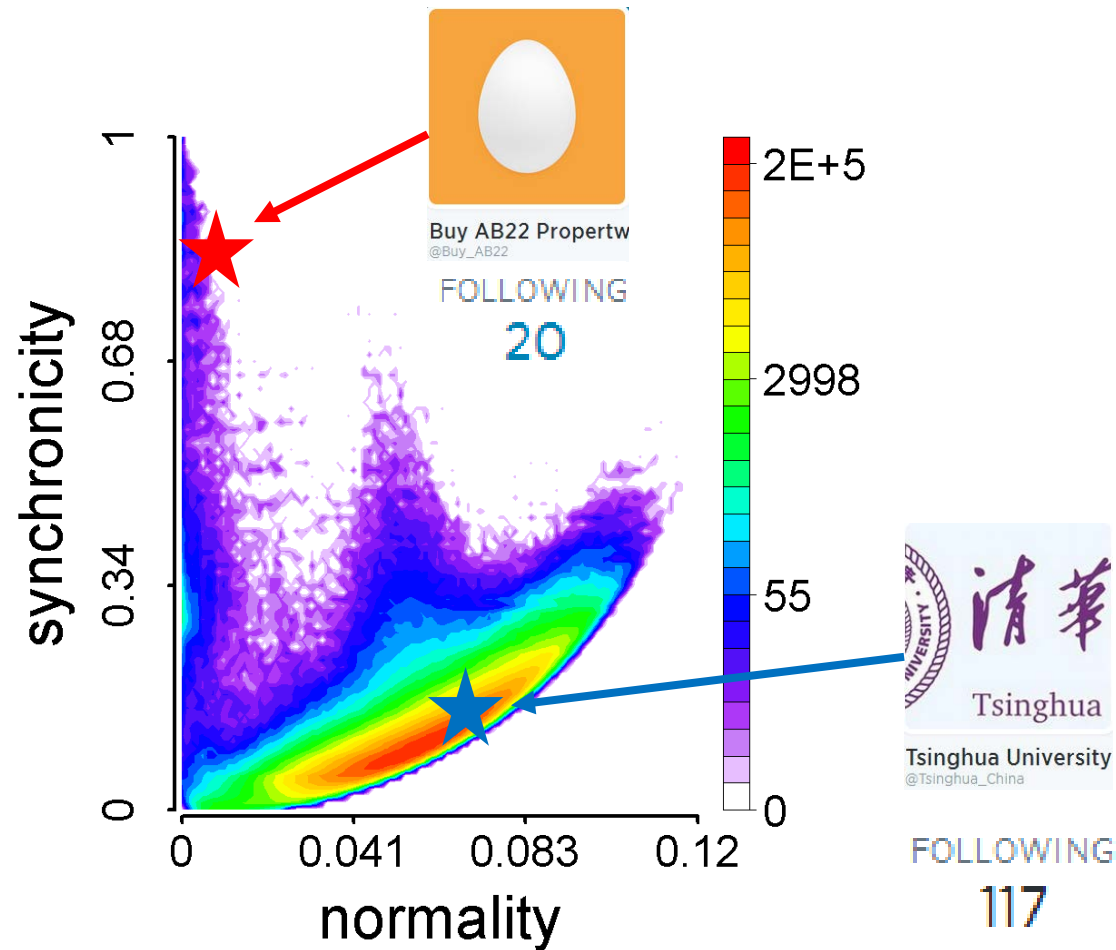
Synchronicity and Normality

- Normality

$$\text{norm}(u) = \frac{\sum_{(v,v') \in \mathcal{F}(u) \times \mathcal{U}} \mathbf{p}(v) \cdot \mathbf{p}(v')}{d(u) \times N}$$



Synchronicity-Normality Plot



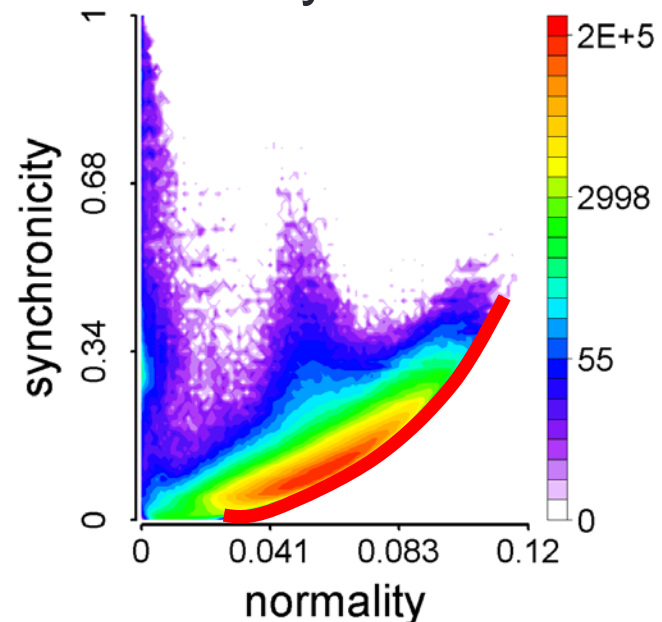
Theorem

- For any distribution, there is a **parabolic lower limit** in the synchronicity-normality plot.

$$s_{min} = (-Mn^2 + 2n - s_b)/(1 - Ms_b)$$

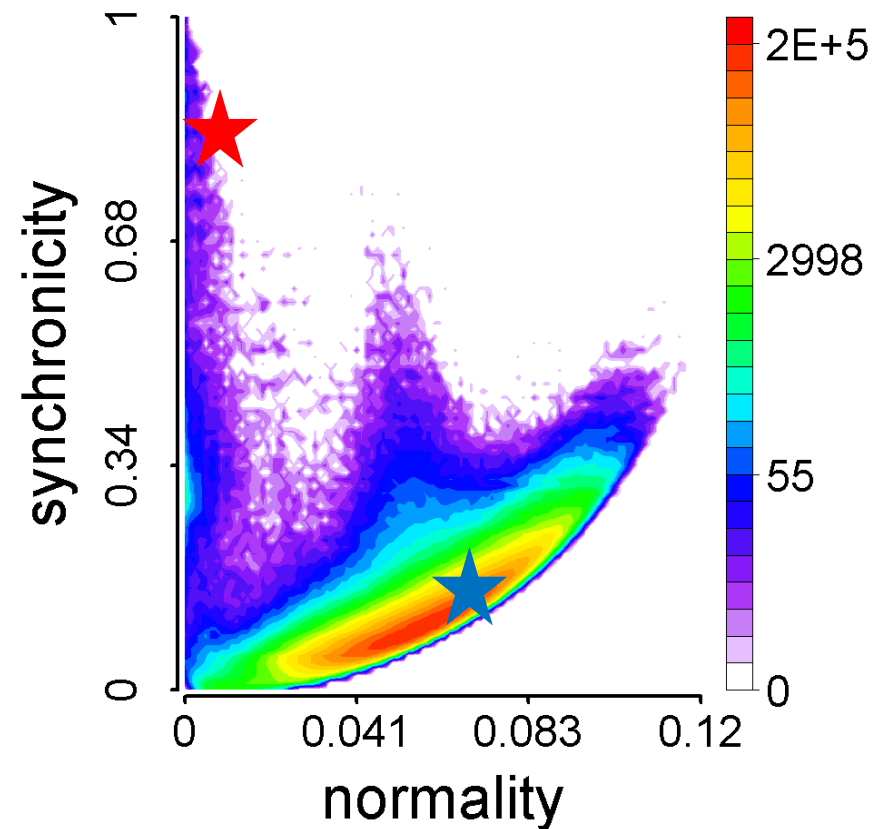
synchronicity \swarrow \nwarrow normality

- Proof. See our paper ☺



CatchSync Algorithm

- Distance-based anomaly detection
- Fraudsters
 - Big synchronicity
 - Small normality
 - Away from the densest



OUTLINE

1. Background

2. Fraudulent Pattern Mining

3. The Algorithm

4. Experiments

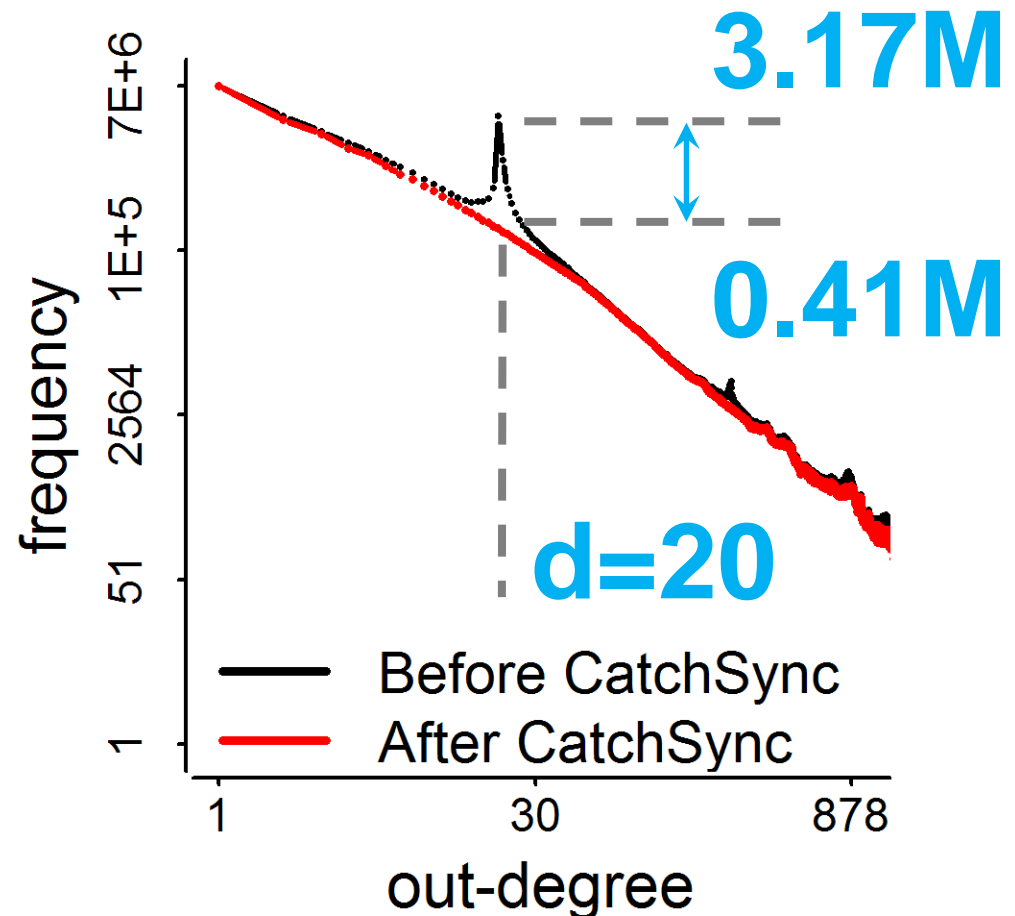
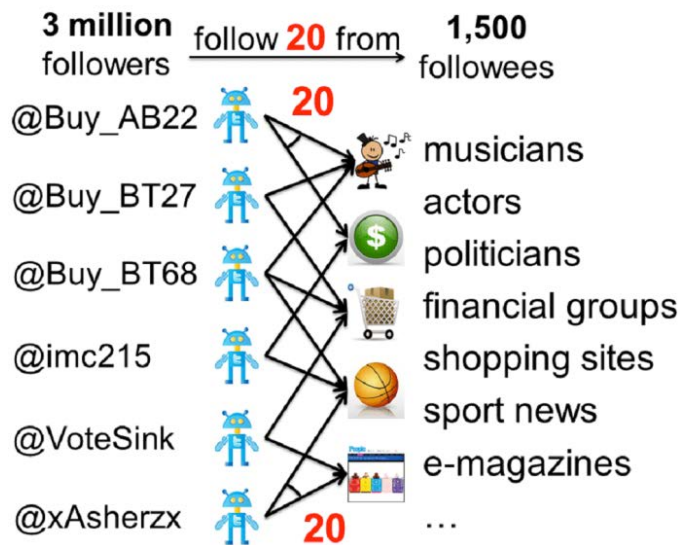
Experiments

- Q1: Does CatchSync remove anomalies?
 - Degree distribution
 - Feature space
- Q2: Is CatchSync catching actually fraudulent users?
- Q3: Is CatchSync robust?

Q1: Does CatchSync Remove Anomalies?



41M

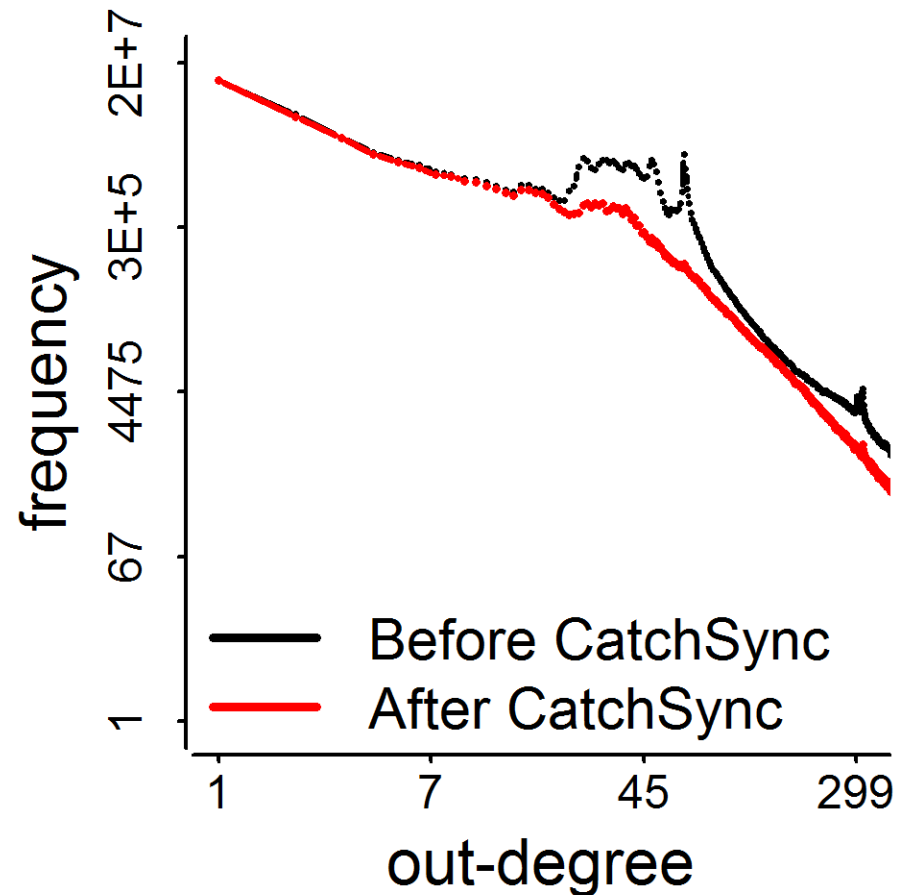


Q1: Does CatchSync Remove Anomalies?

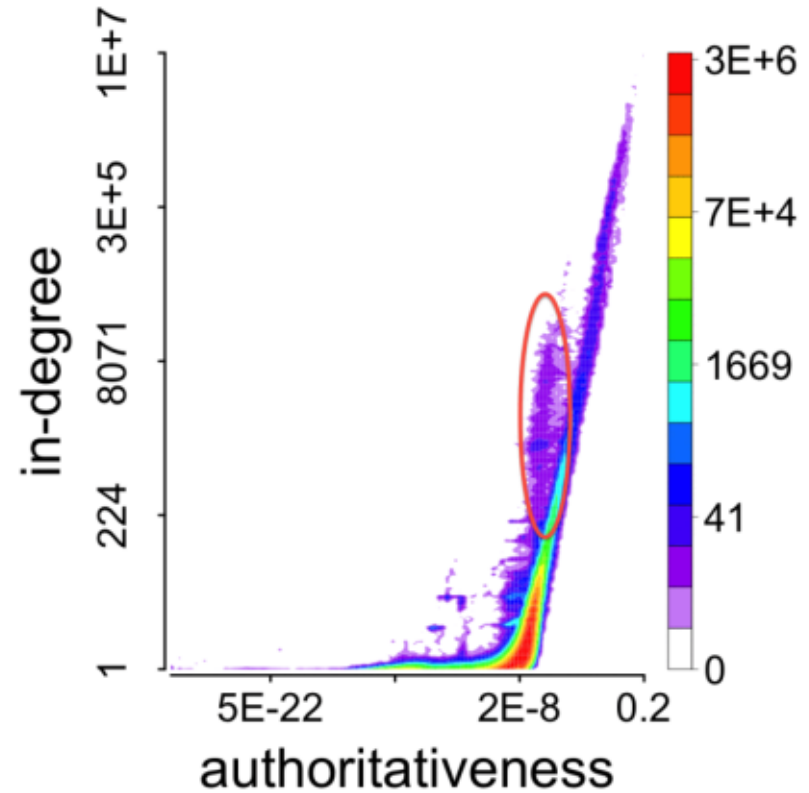
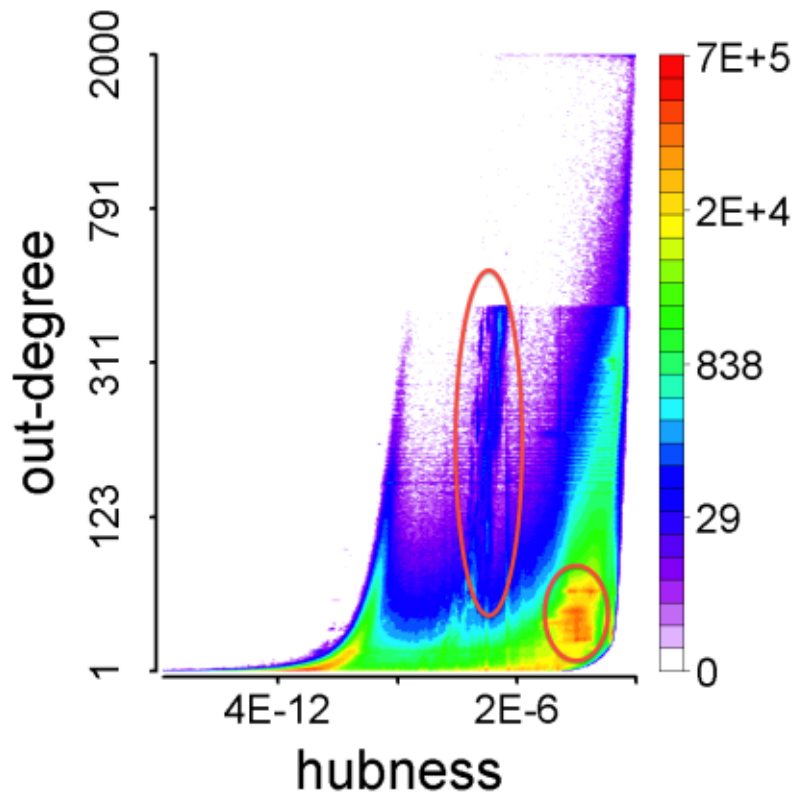
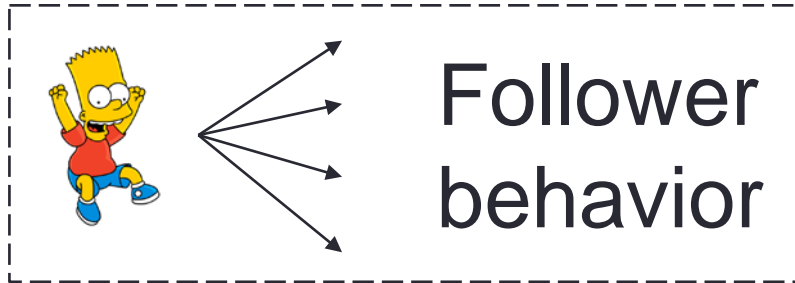


2011

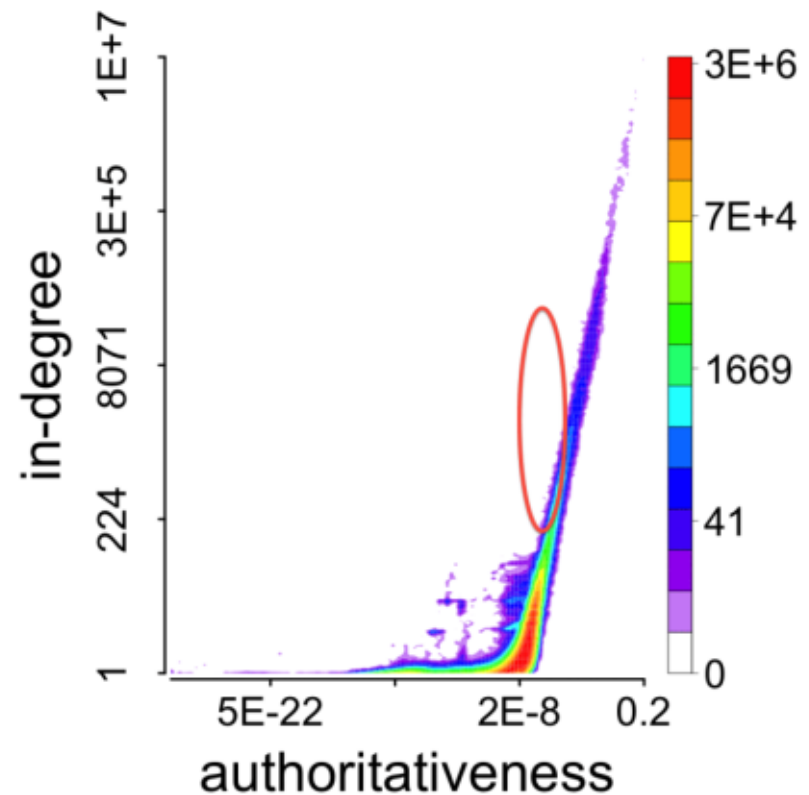
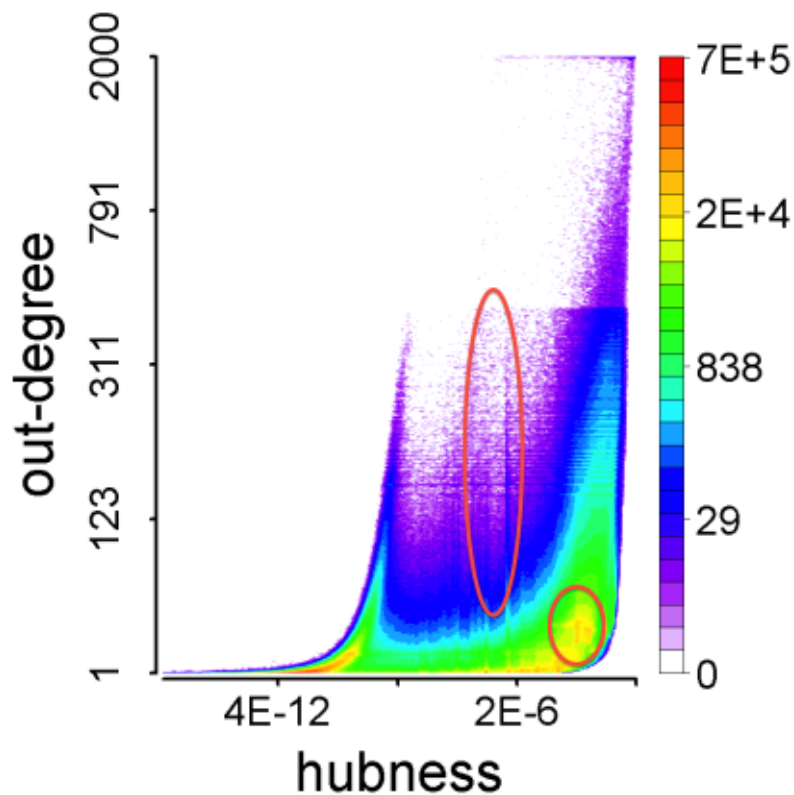
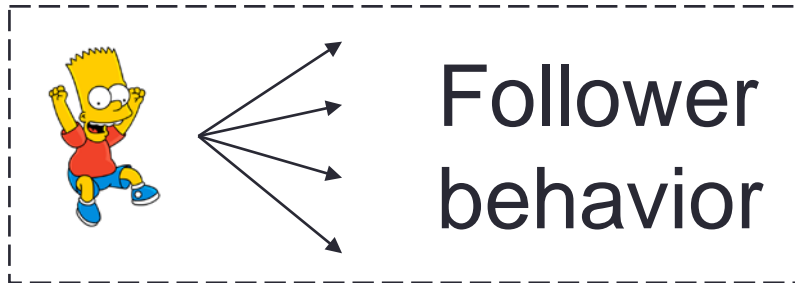
117M



Before CatchSync



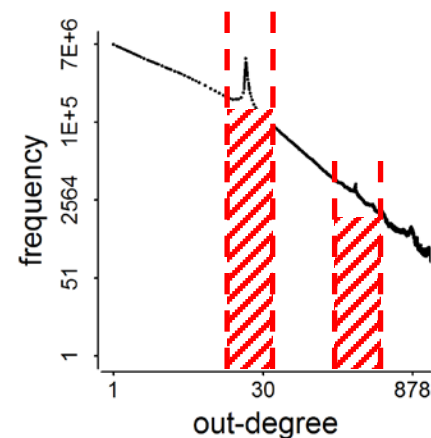
After CatchSync



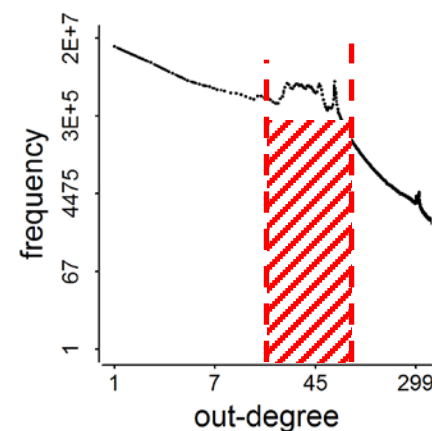
Q2: Is CatchSync Catching Actually Fraudulent Users?



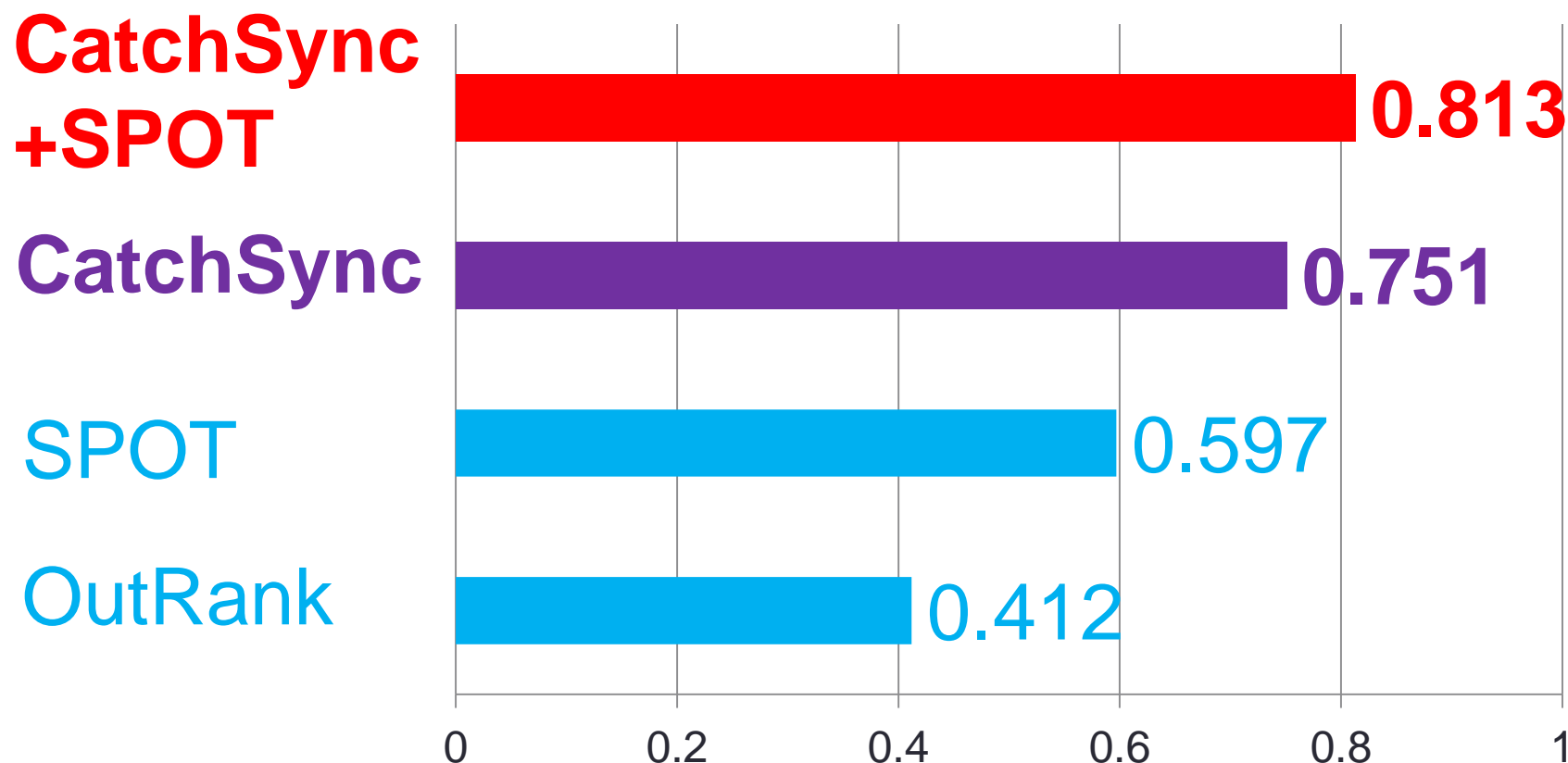
173/1,000



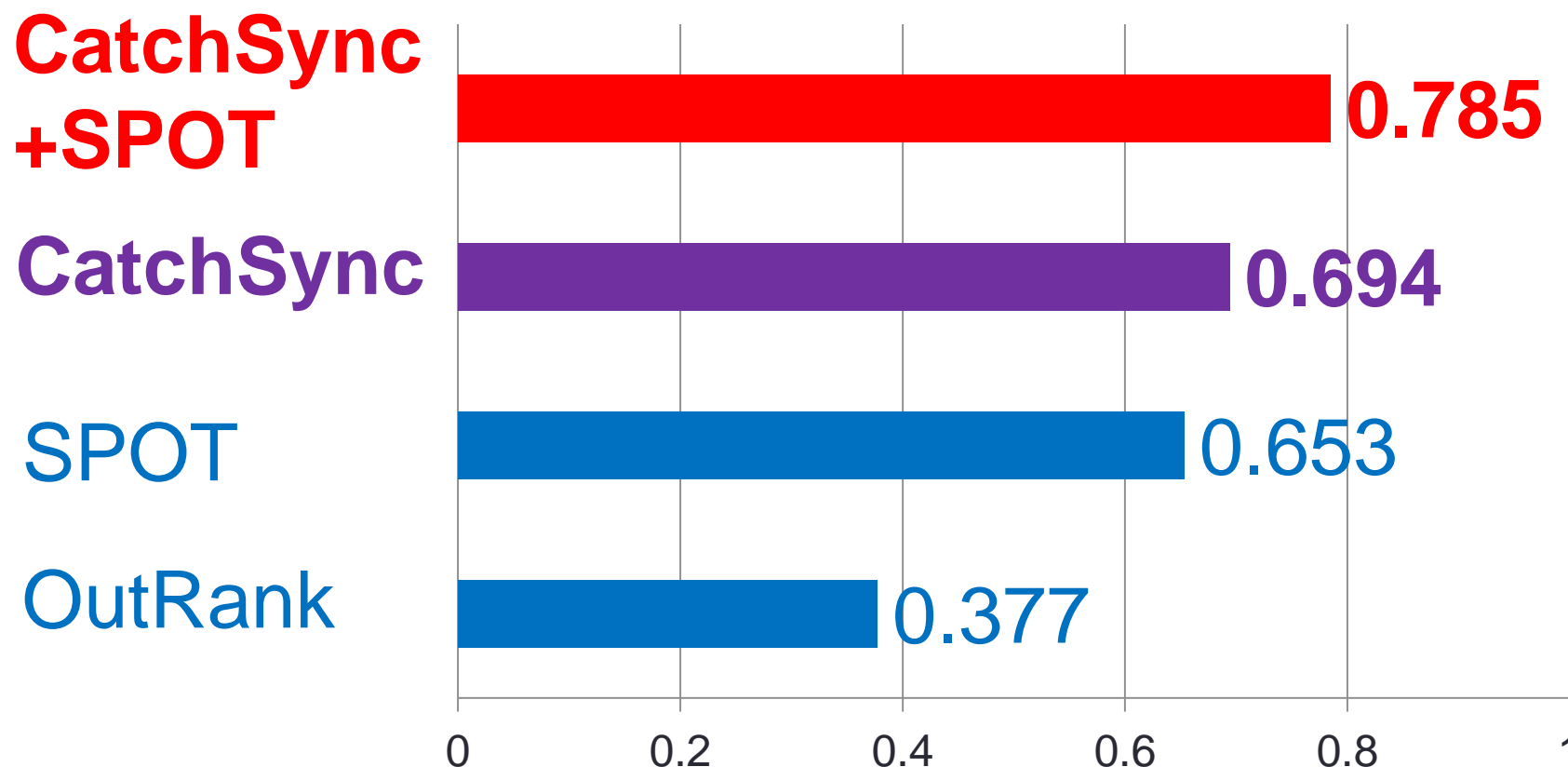
237/1,000



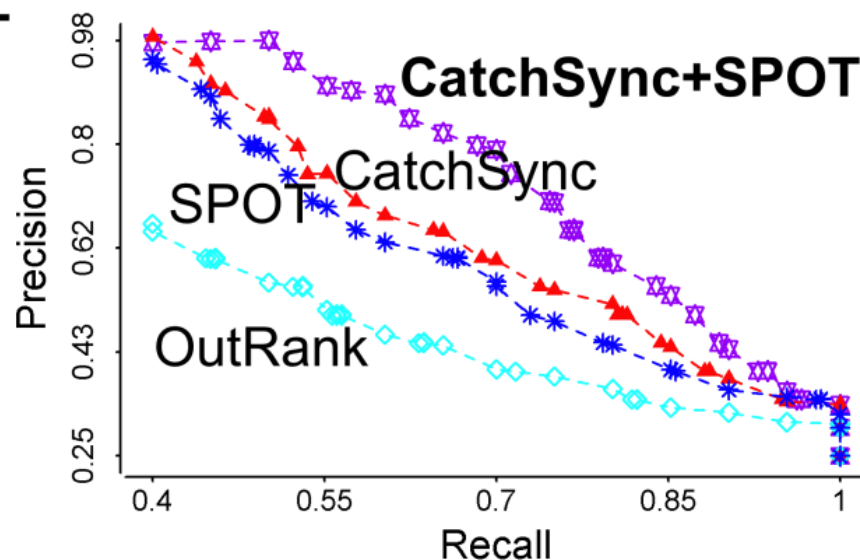
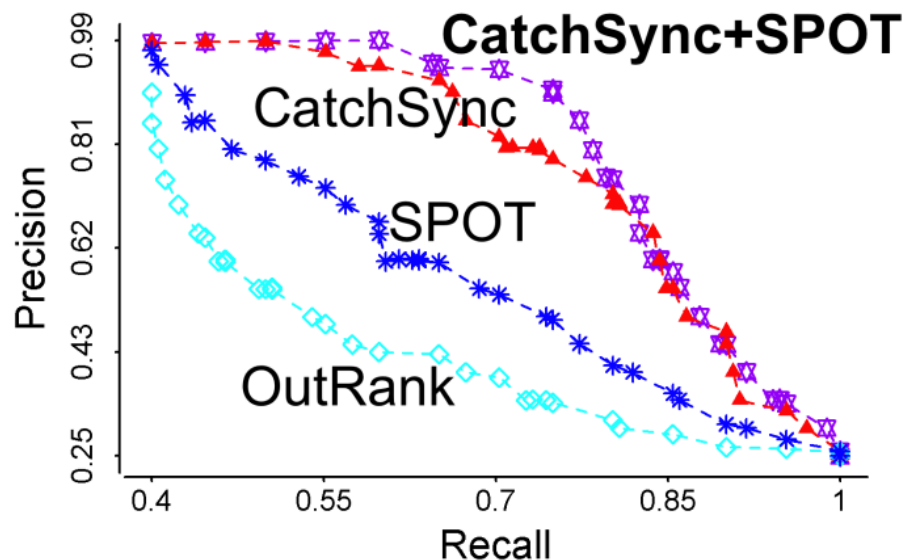
Q2: Is CatchSync Catching Actually Fraudulent Users?



Q2: Is CatchSync Catching Actually Fraudulent Users?



Q2: Is CatchSync Catching Actually Fraudulent Users?



Recall = 80%

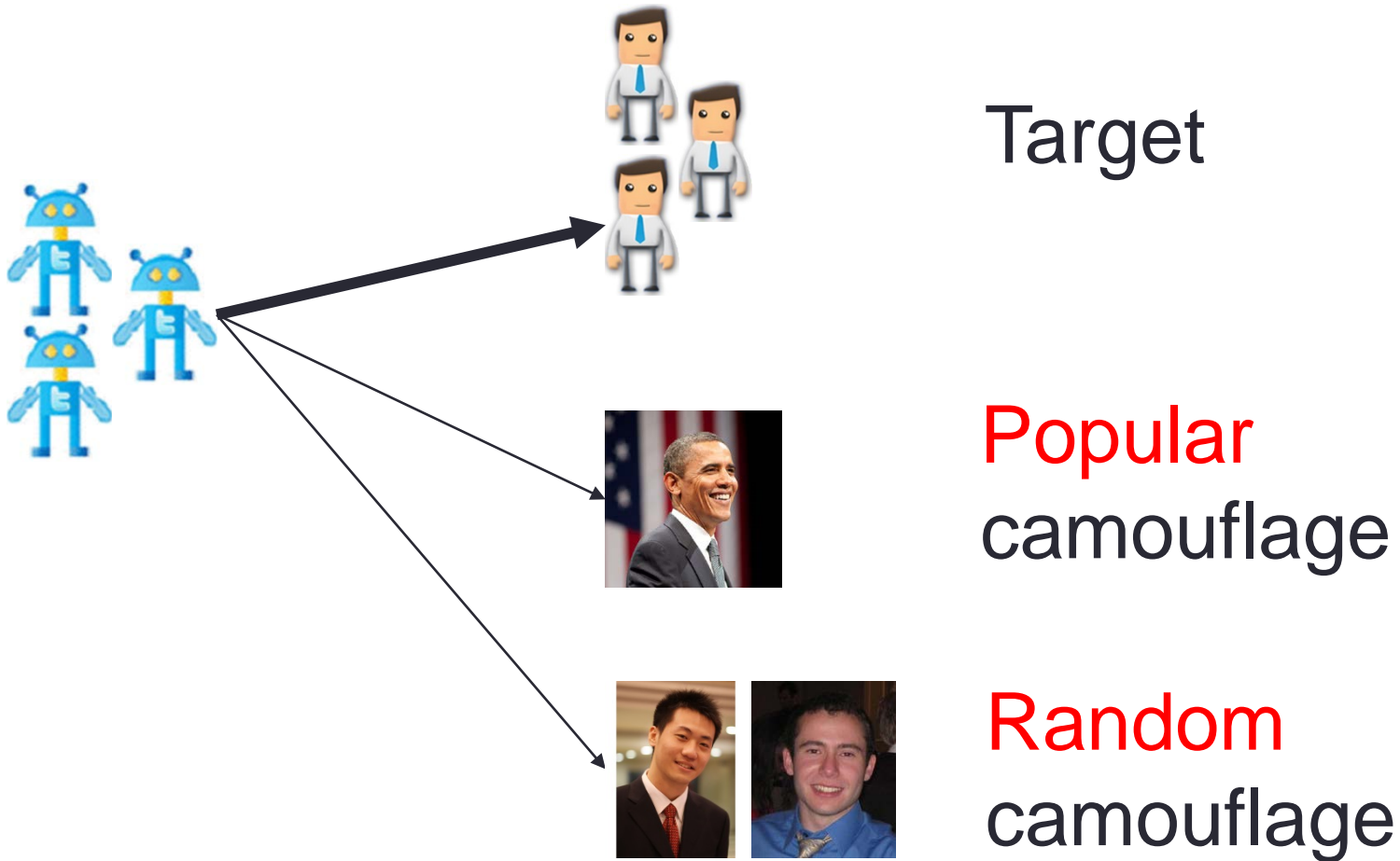
Precision in Twitter

Precision in Tencent Weibo

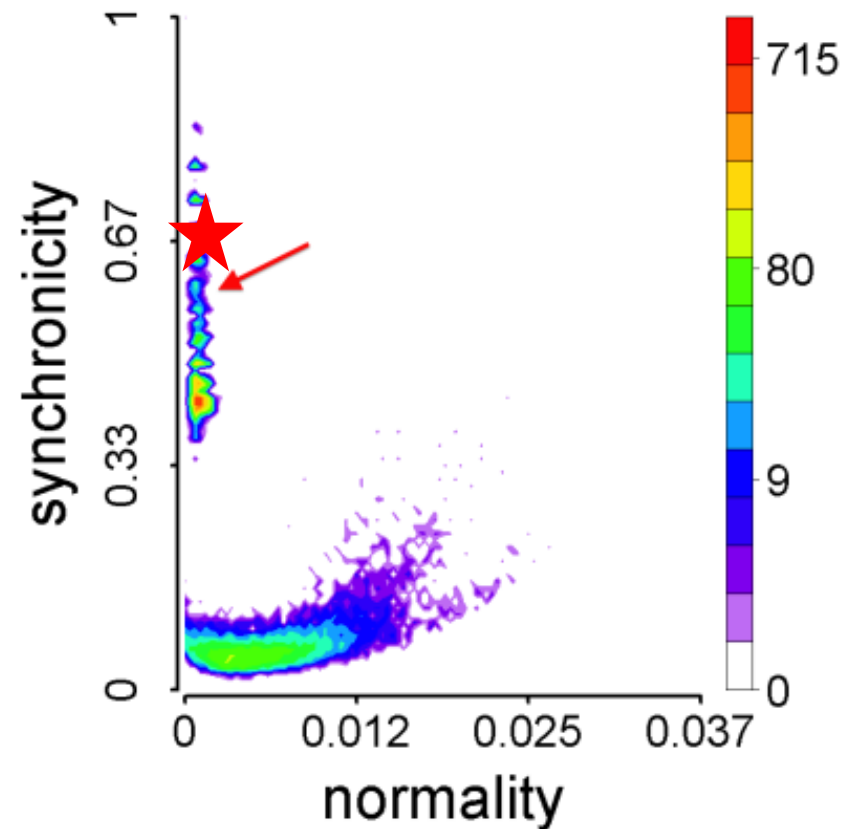
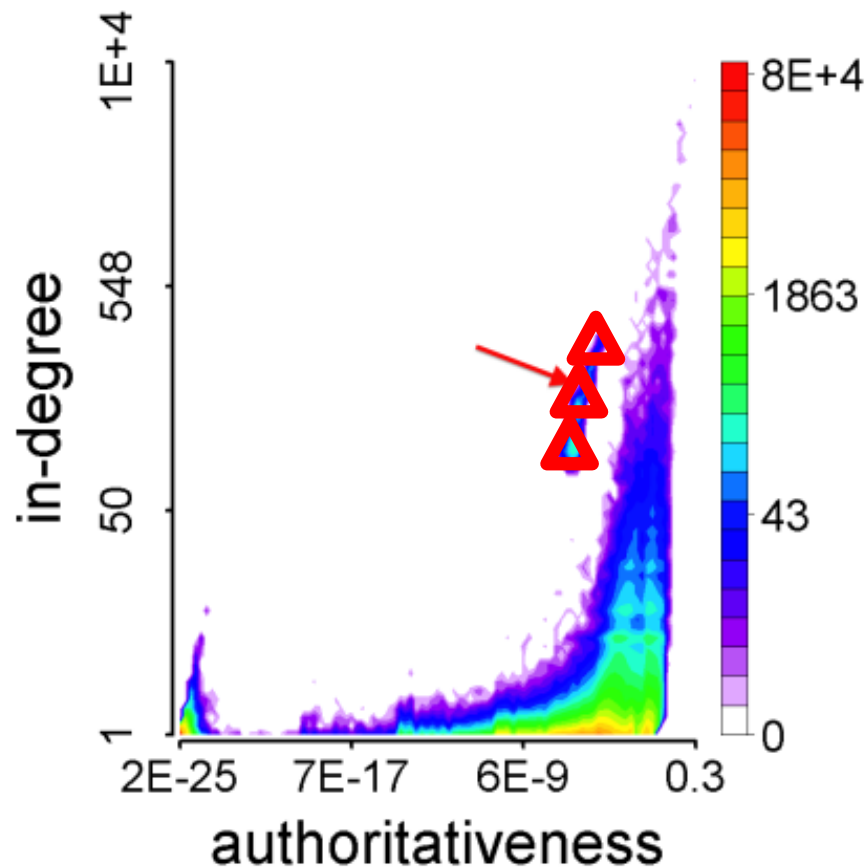
83.5%

79.4%

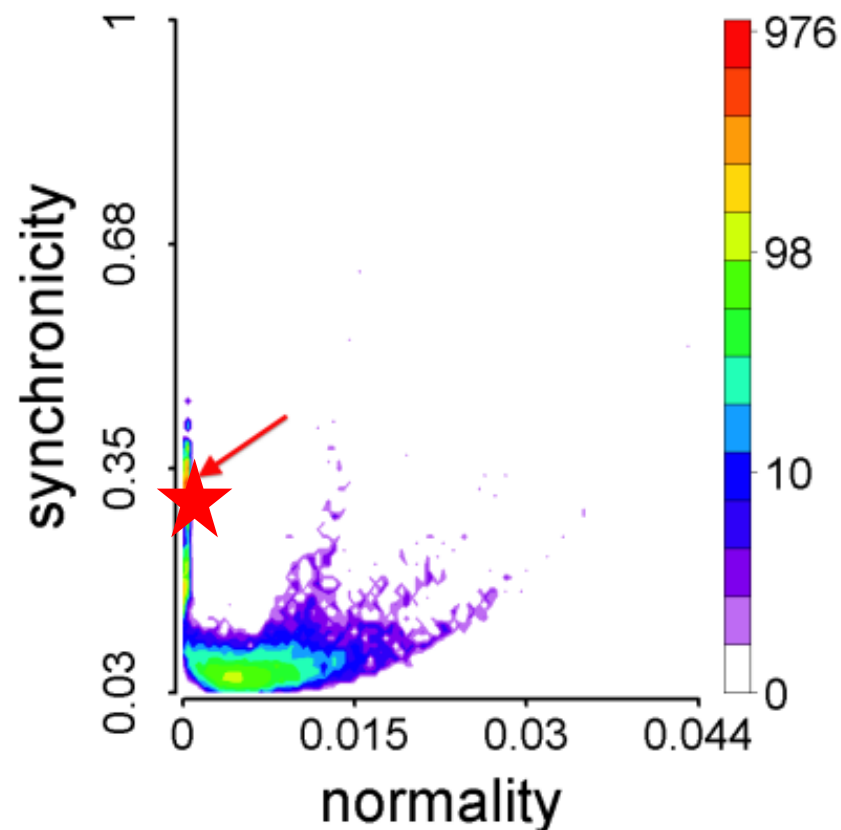
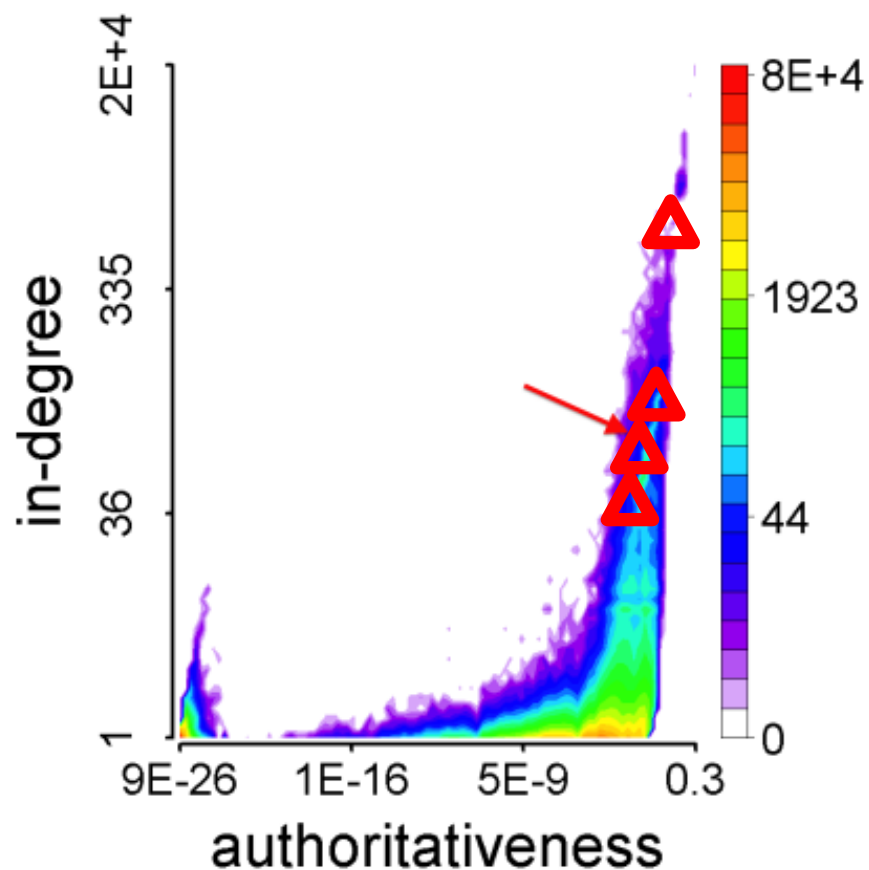
Q3: Is CatchSync Robust to Camouflage?



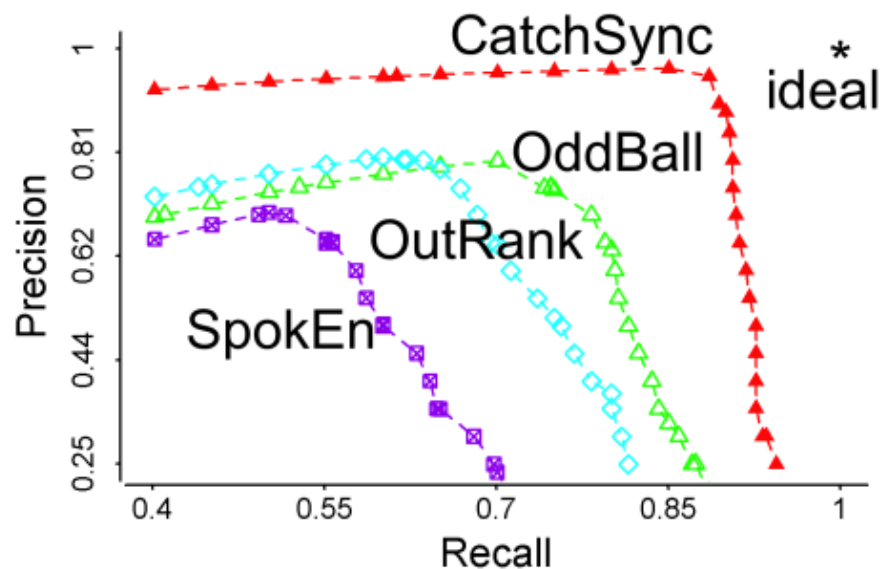
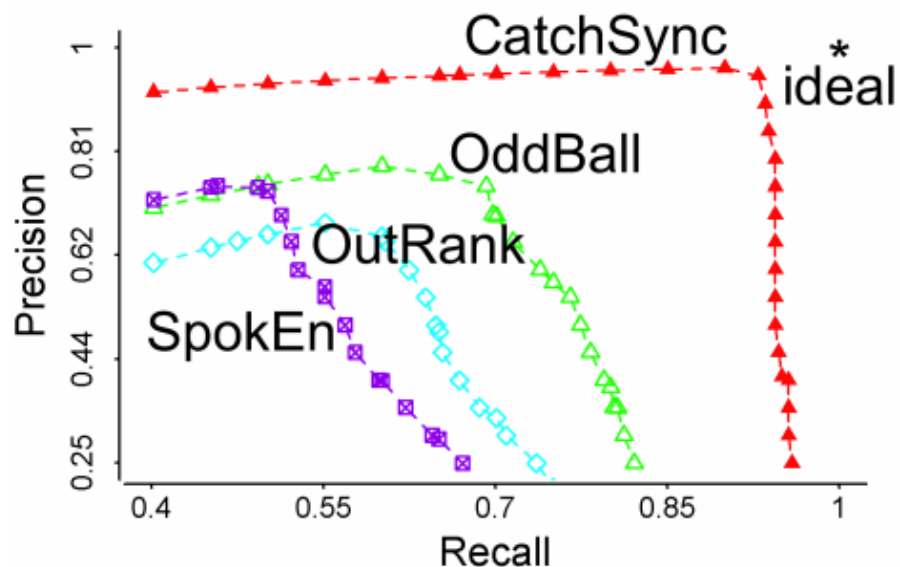
Q3: Is CatchSync Robust to Camouflage?



Q3: Is CatchSync Robust to Camouflage?



Q3: Is CatchSync Robust to Camouflage?



Conclusion

- Goals
 - G1. Find **patterns** that **distinguish** fraudulent user behavior from normal behavior
 - **A1: Synchronized & Abnormal!**
 - G2. Design **algorithms** that **catch** fraudsters
 - **A2: CatchSync!**
 - Remove spikes
 - Content free
 - Robust to camouflage

Questions?

Meng Jiang
 mjiang89@gmail.com
<http://www.meng-jiang.com>

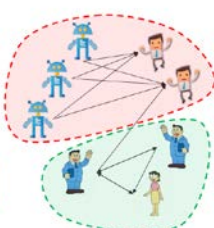


Fraud Detection: Graph Analysis Problem

Twitter

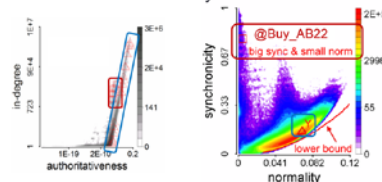


Facebook



CatchSync Algorithm

- Fraudsters
 - Big synchronicity and small normality
 - Away from the densest
- Distance-based anomaly detection method



Q1: Does CatchSync Catch the Anomalies?

Twitter 2009

