BEHAVIORAL MODELING IN SOCIAL NETWORKS FROM MICRO TO MACRO

Meng Jiang, *University of Illinois at Urbana-Champaign*
Peng Cui, *Tsinghua University*

ICDM 2015 TUTORIAL
Atlantic City, NJ

清華大學 Tsinghua University

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Thanks to

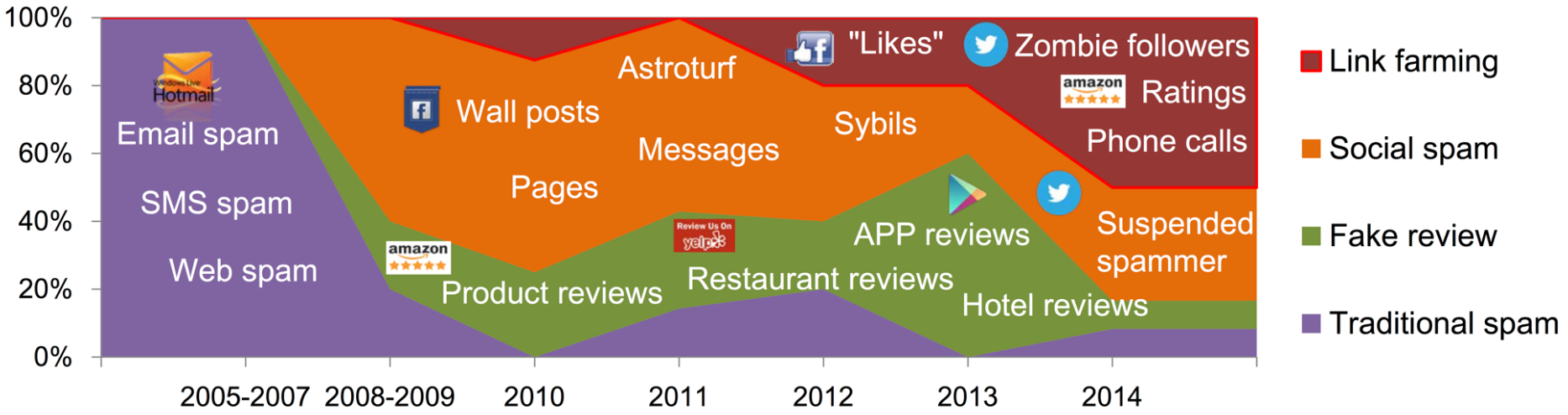
TSINGHUA UNIVERSITY- TENCENT
JOINT LABORATORY

National Natural Science Foundation of China,
No. 61370022, No. 61210008, No. 61303075
International Science & Technology Cooperation Program of China,
No. 2013DFG12870
National Program on Key Basic Research Project,
No. 2011CB302206, No. 2015CB352300
NExT Research Center funded by MDA, Singapore,
WBS:R-252-300-001-490
Tsinghua-Tencent Joint Laboratory

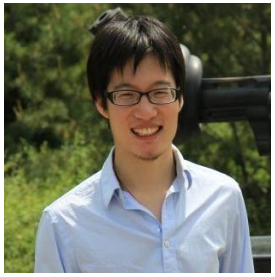
Outline

- Prediction for natural behavior
  - Modeling individual behavior (MICRO)
  - Modeling information cascade (MACRO)
- **Detection for unnatural behavior**
  - **Suspicious behavior detection**

Suspicious Behavior Detection



JIANG, Meng (UIUC)



CUI, Peng (Tsinghua)

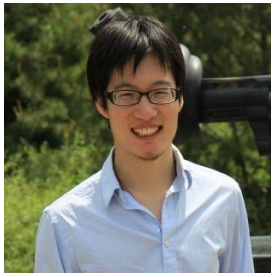
FALOUTSOS, Christos (CMU)

Suspicious Behavior Detection: Current Trends and Future Directions.

Special Issue on Online Behavioral Analysis and Modeling, IEEE Intelligent Systems Magazine (ISSI), 2016. (to appear)

Suspicious Behavior Detection

JIANG, Meng (UIUC)



CUI, Peng (Tsinghua)

FALOUTSOS, Christos (CMU)

Suspicious Behavior Detection: Current Trends and Future Directions.

Special Issue on Online Behavioral Analysis and Modeling, IEEE Intelligent Systems Magazine (ISSI), 2016. (to appear)

Suspicious Behaviors

- Selling Twitter followers

Suspicious Behaviors

❖Selling Facebook Likes

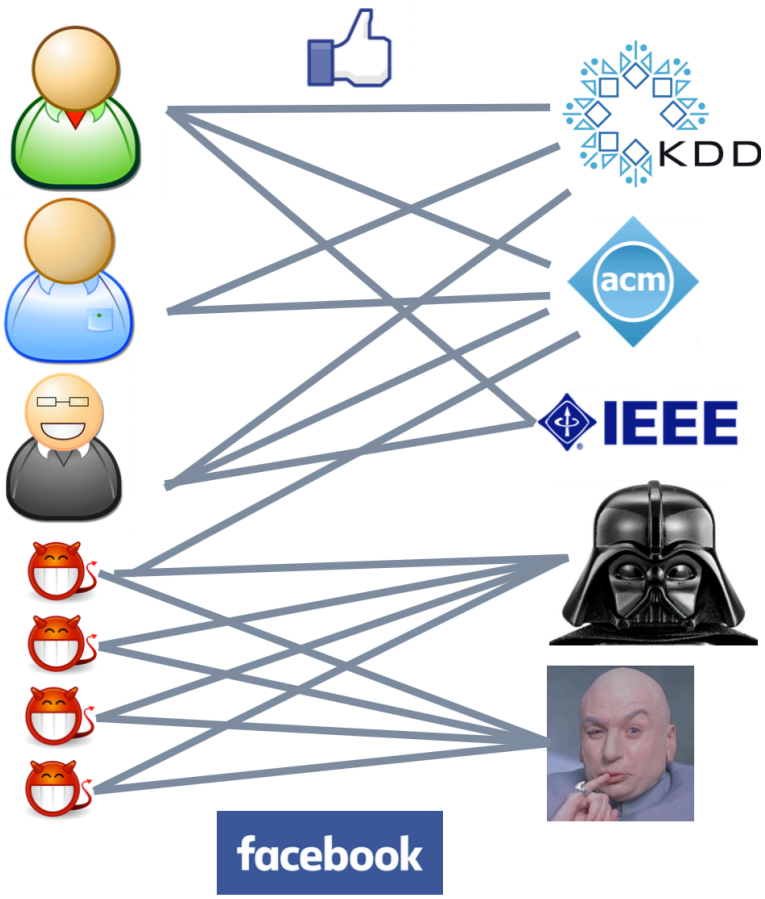
Suspicious Behavior Detection

- Detecting suspicious behavioral patterns
  - **Spotting:** Lockstep patterns
    - Methods that can spot strange behaviors
  - **Catching:** Synchronized patterns
    - Scalable algorithms with theoretical guarantee
  - **Solving:** Suspiciousness in multiple dimensions
    - A principled metric for suspiciousness

Lockstep Behavior: Facebook Likes

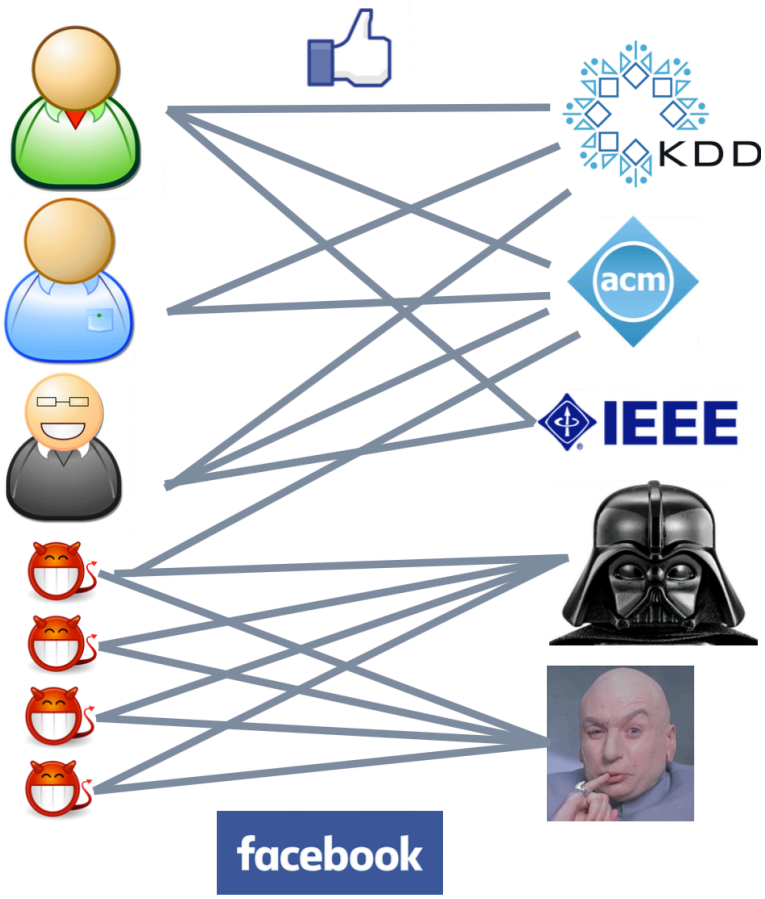
Beutel et al. **CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Neworks.** *WWW*, 2013.

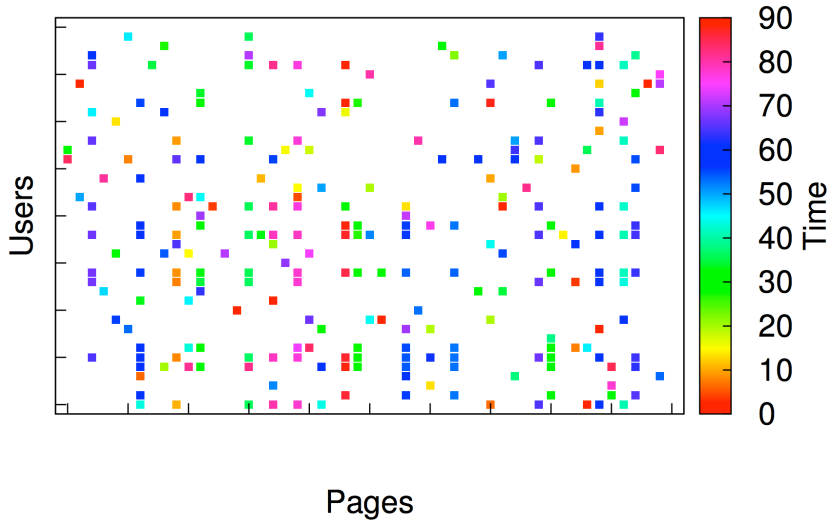
Lockstep Behavior: Facebook Likes



Beutel et al. **CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Neworks.** *WWW*, 2013.

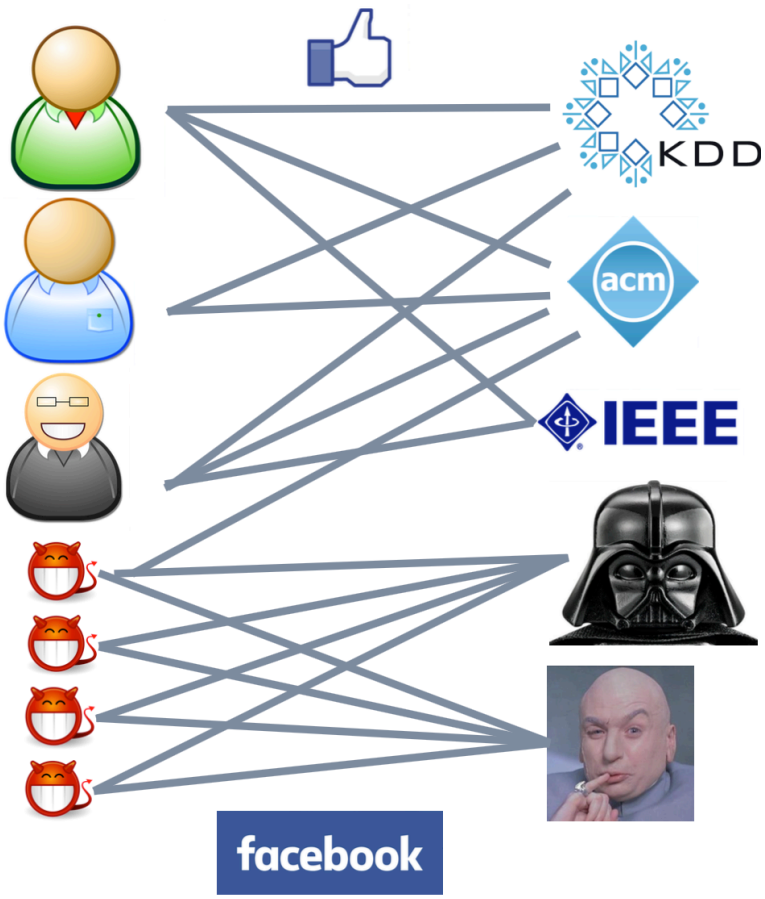
Lockstep Behavior: Graphical View

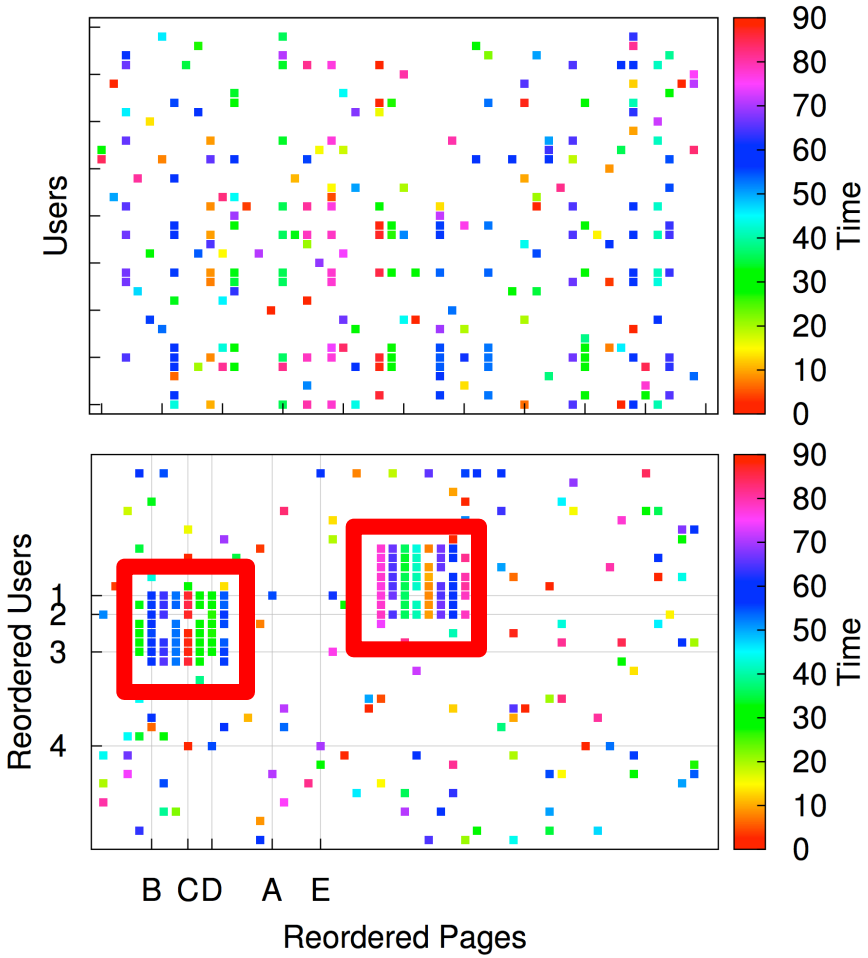




Beutel et al. **CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Neworks.** *WWW*, 2013.

Lockstep Behavior: Reorder Matrix

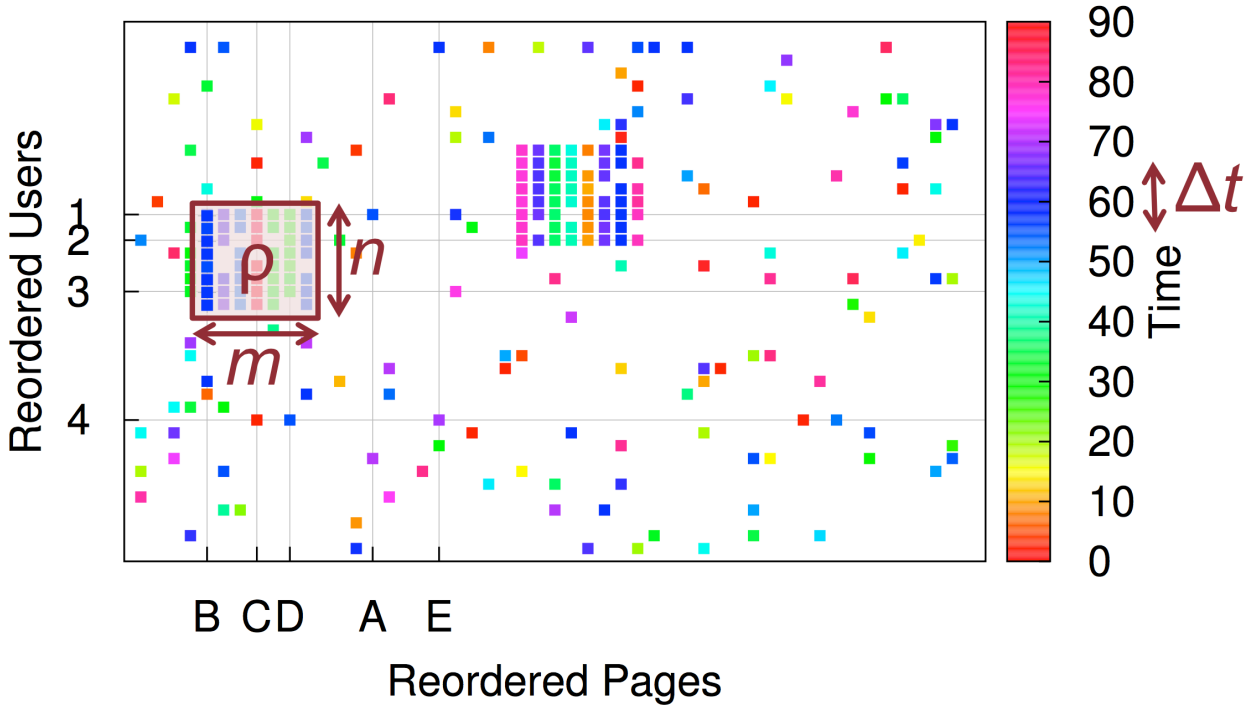




Beutel et al. **CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Neworks.** *WWW*, 2013.

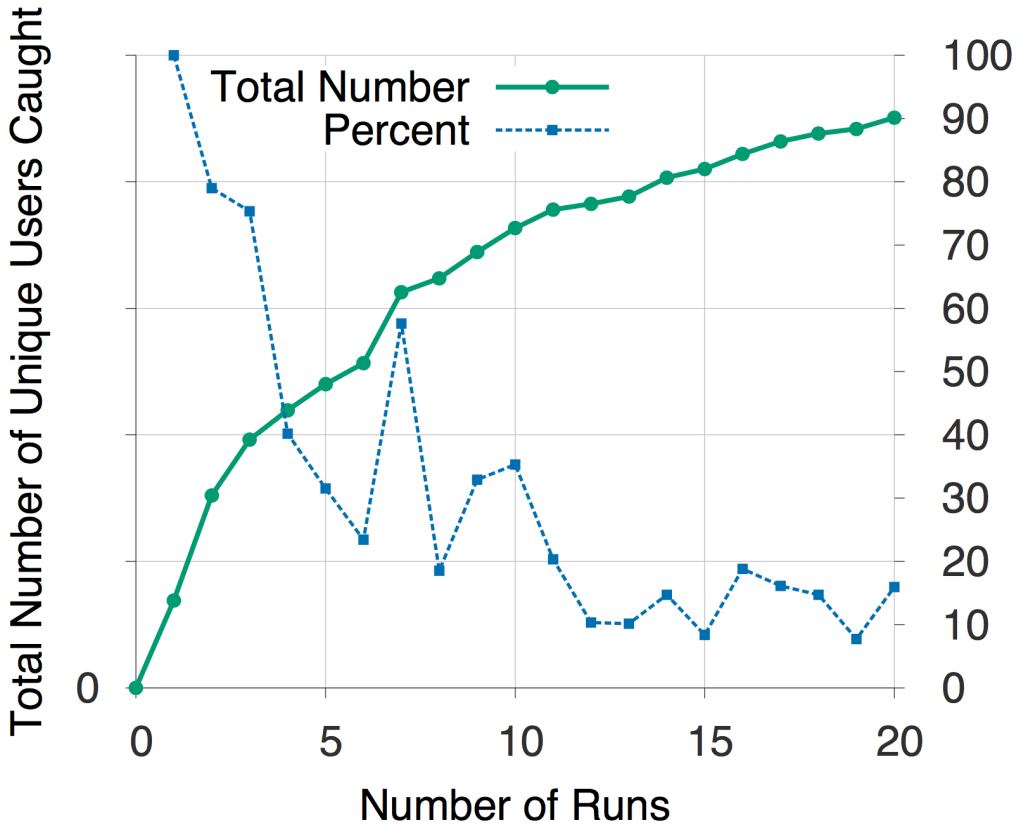
Lockstep Behavior: Seed + Search

- CopyCatch
  - "Near Bipartite Core": *n* users, *m* Pages, *ρ*, *Δt*



Lockstep Behavior: Performance

- CopyCatch

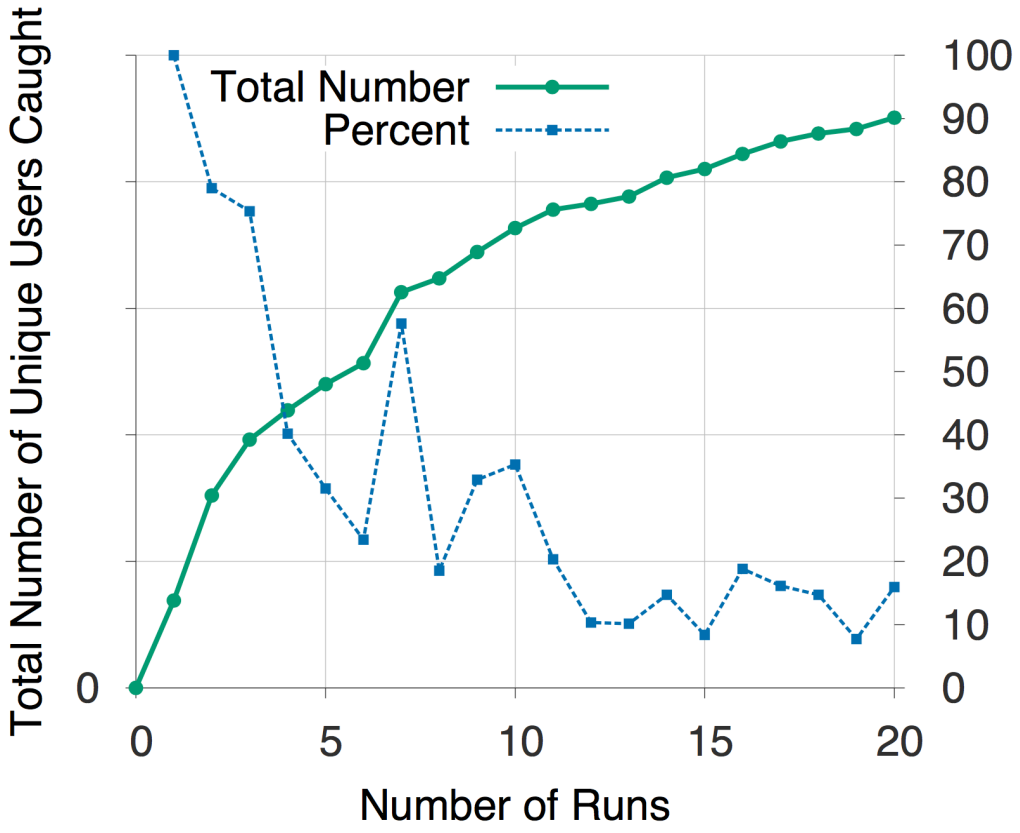


Beutel et al. **CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Neworks.** *WWW*, 2013.

Lockstep Behavior: Performance

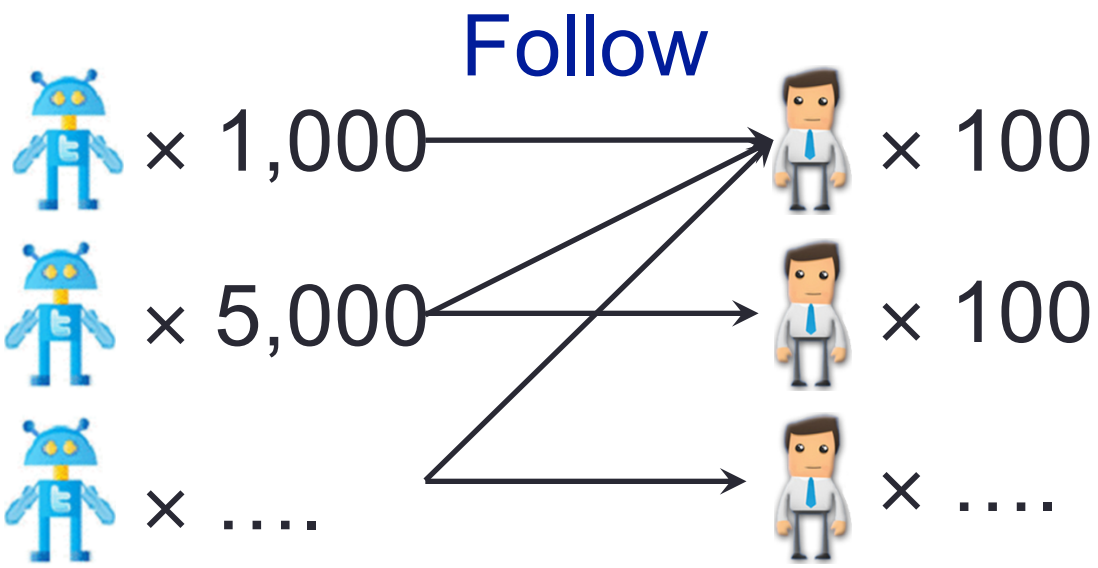
- CopyCatch

Seed Selection!!!



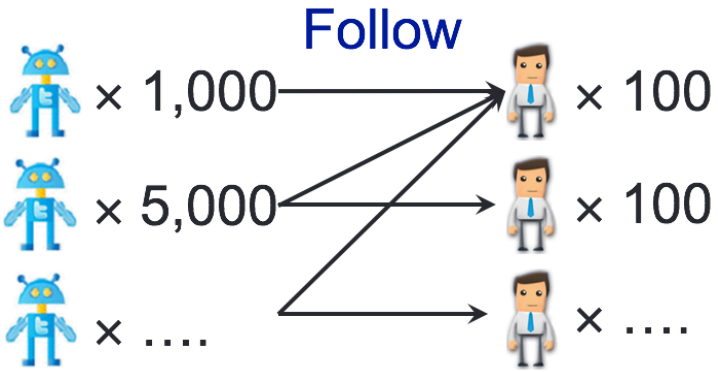
Beutel et al. **CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Neworks.** *WWW*, 2013.

Lockstep Behavior: Twitter Followers

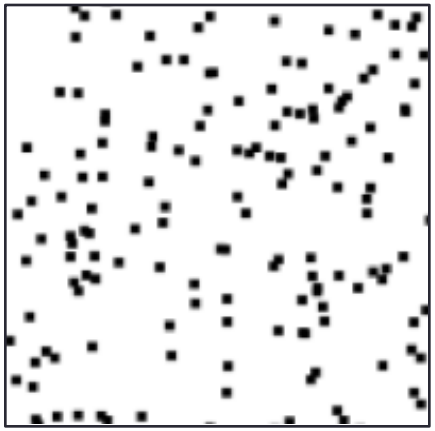


Jiang et al. **Inferring Strange Behavior from Connectivity Pattern in Social Networks.** *PAKDD*, 2014.

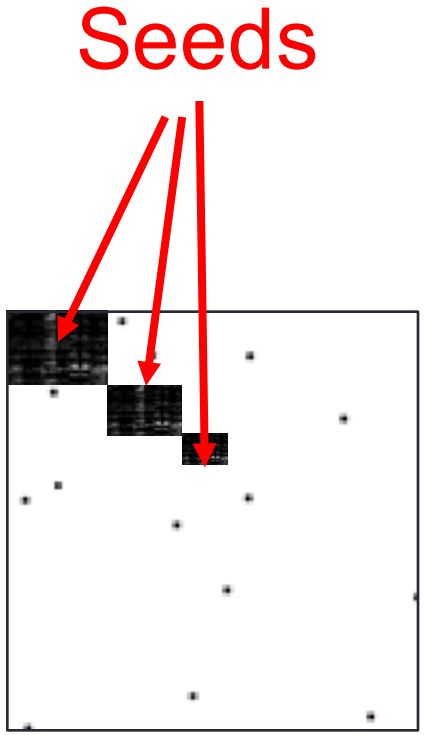
Lockstep Behavior: Reorder Matrix



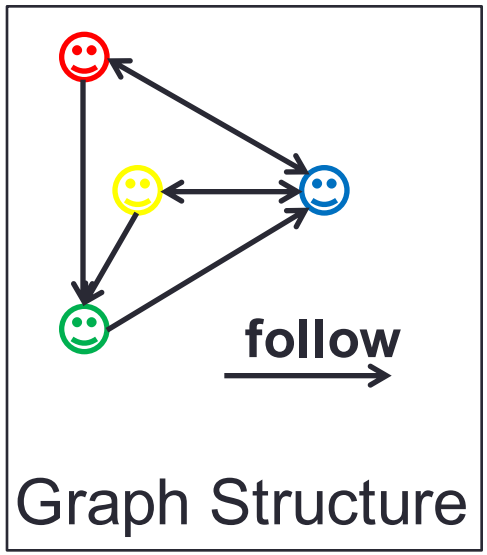
Given

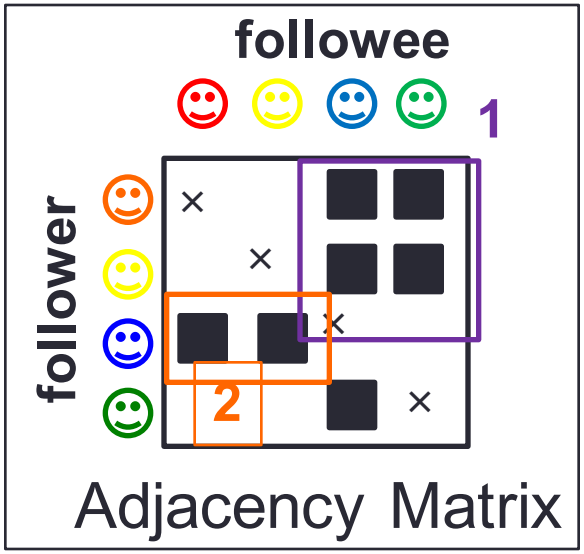


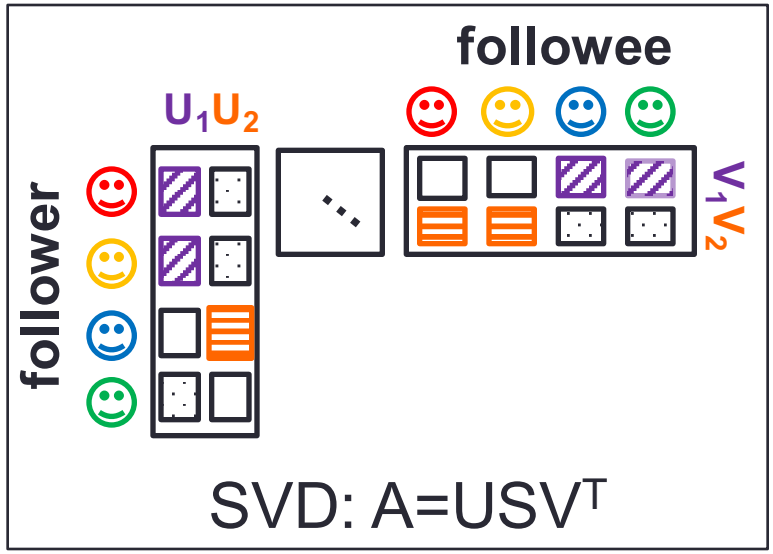
Reorder



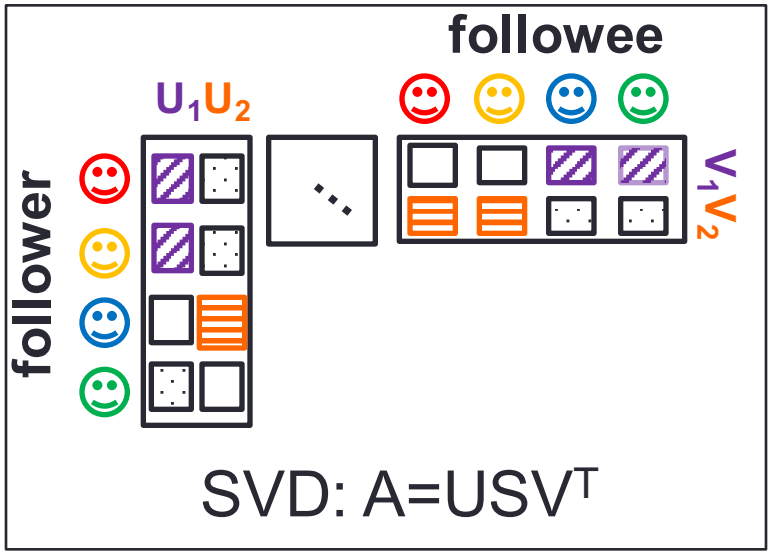
Lockstep Behavior: SVD Reminder

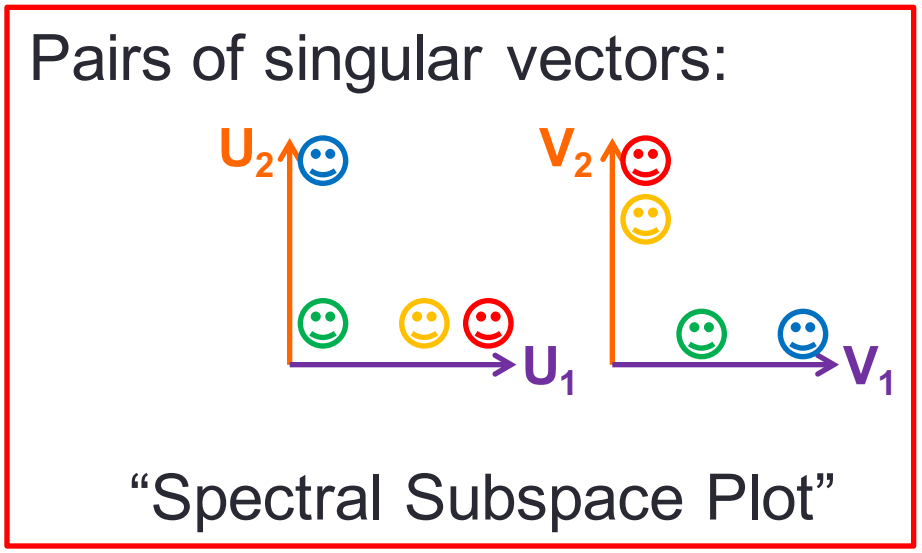






Lockstep Behavior: Spectral Subspace

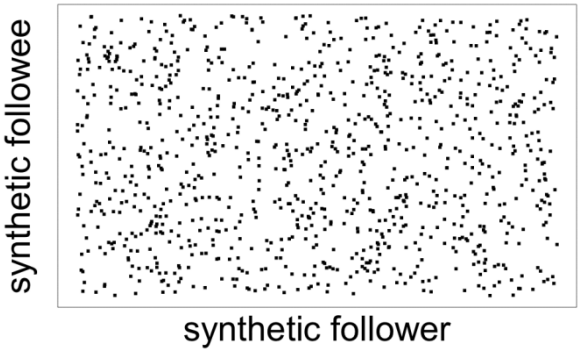




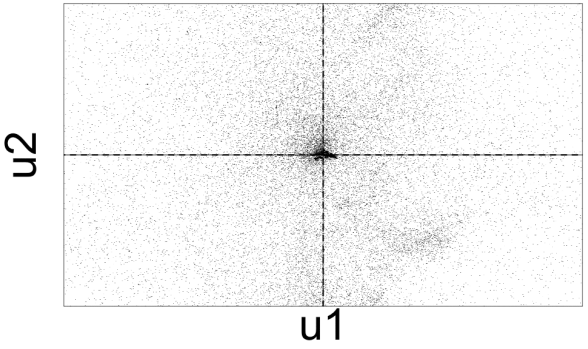
Spectral Subspace Plot: Case #0

- NO lockstep behavior: Scatter

Adjacency Matrix



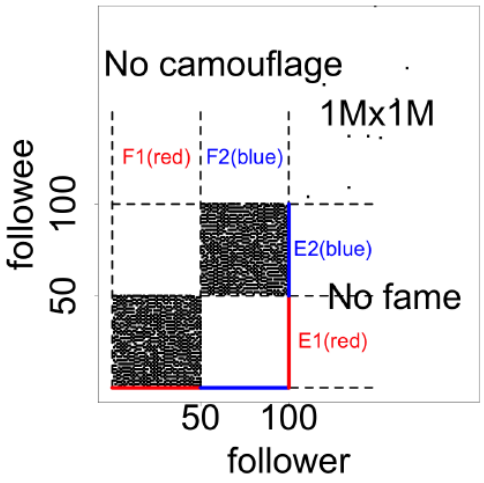
Spectral Subspace Plot



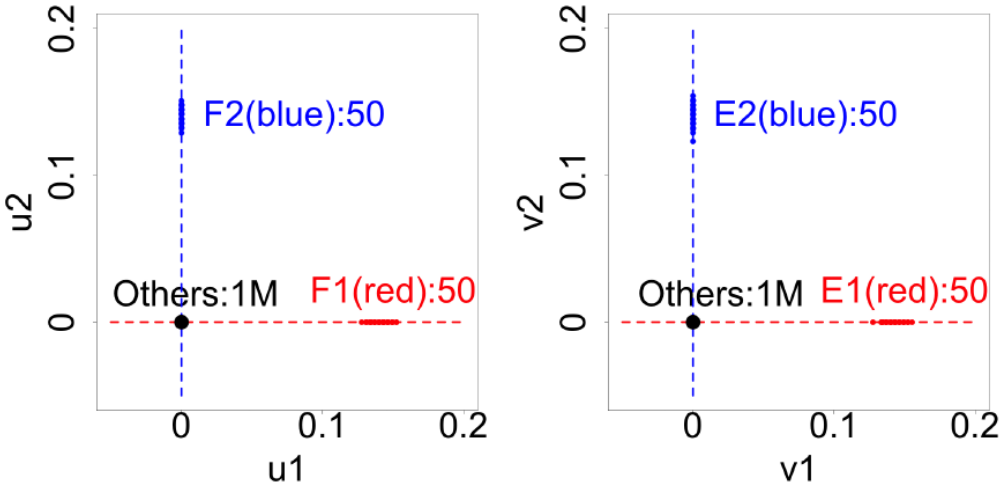
Spectral Subspace Plot: Case #1

- Non-overlapping lockstep: "Rays"

Adjacency Matrix



Spectral Subspace Plot

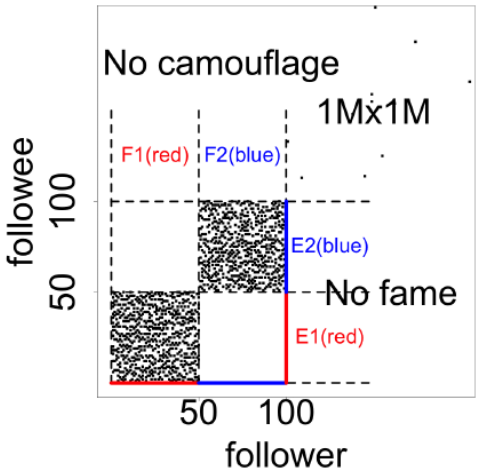


Rule 1 (short "rays"): two blocks, high density (90%), no "camouflage", no "fame"

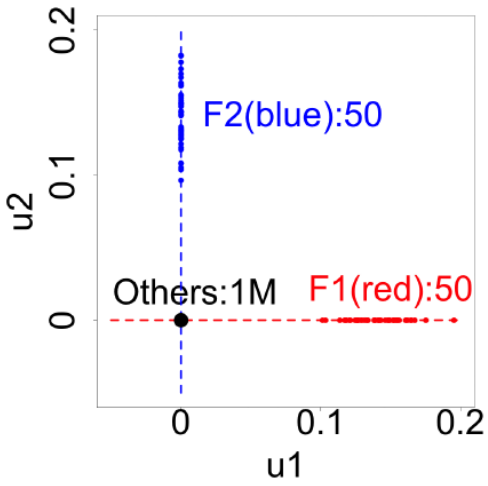
Spectral Subspace Plot: Case #2

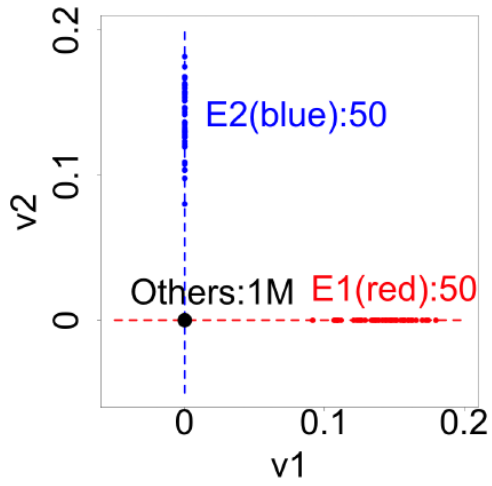
- Non-overlapping: Low density, Elongation

Adjacency Matrix



Spectral Subspace Plot



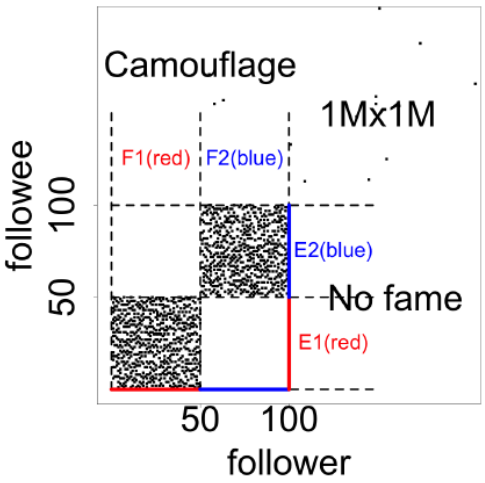


Rule 2 (long "rays"): two blocks, low density (50%), no "camouflage", no "fame"

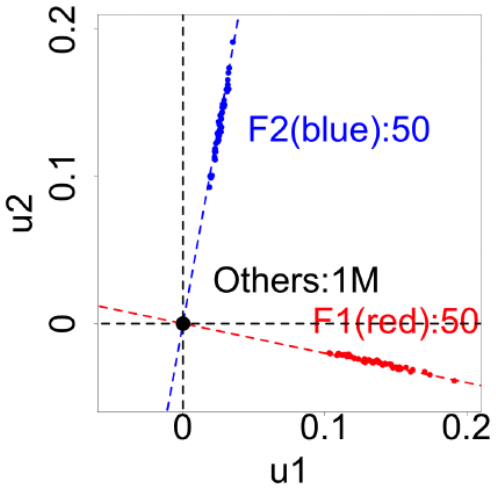
Spectral Subspace Plot: Case #3

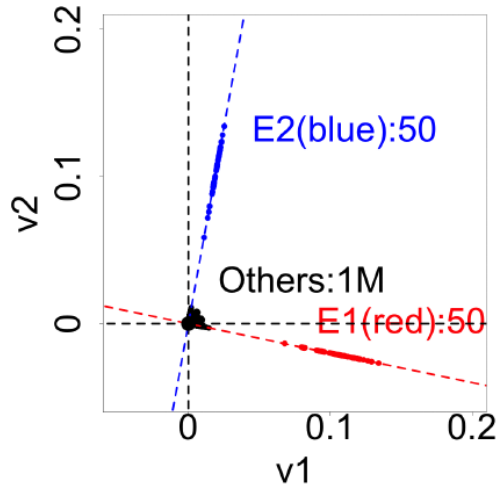
- Non-overlapping: Camouflage/Fame, Tilting

Adjacency Matrix



Spectral Subspace Plot

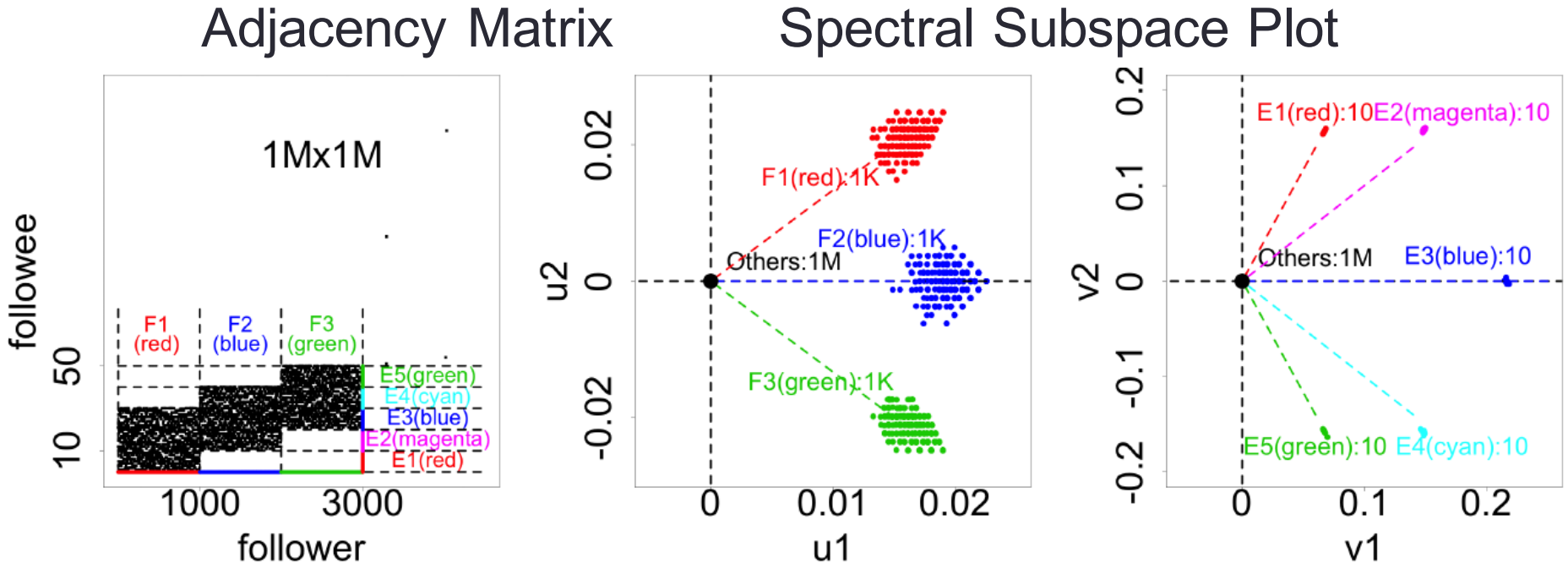




Rule 3 (tilting "rays"): two blocks, with "camouflage", no "fame"

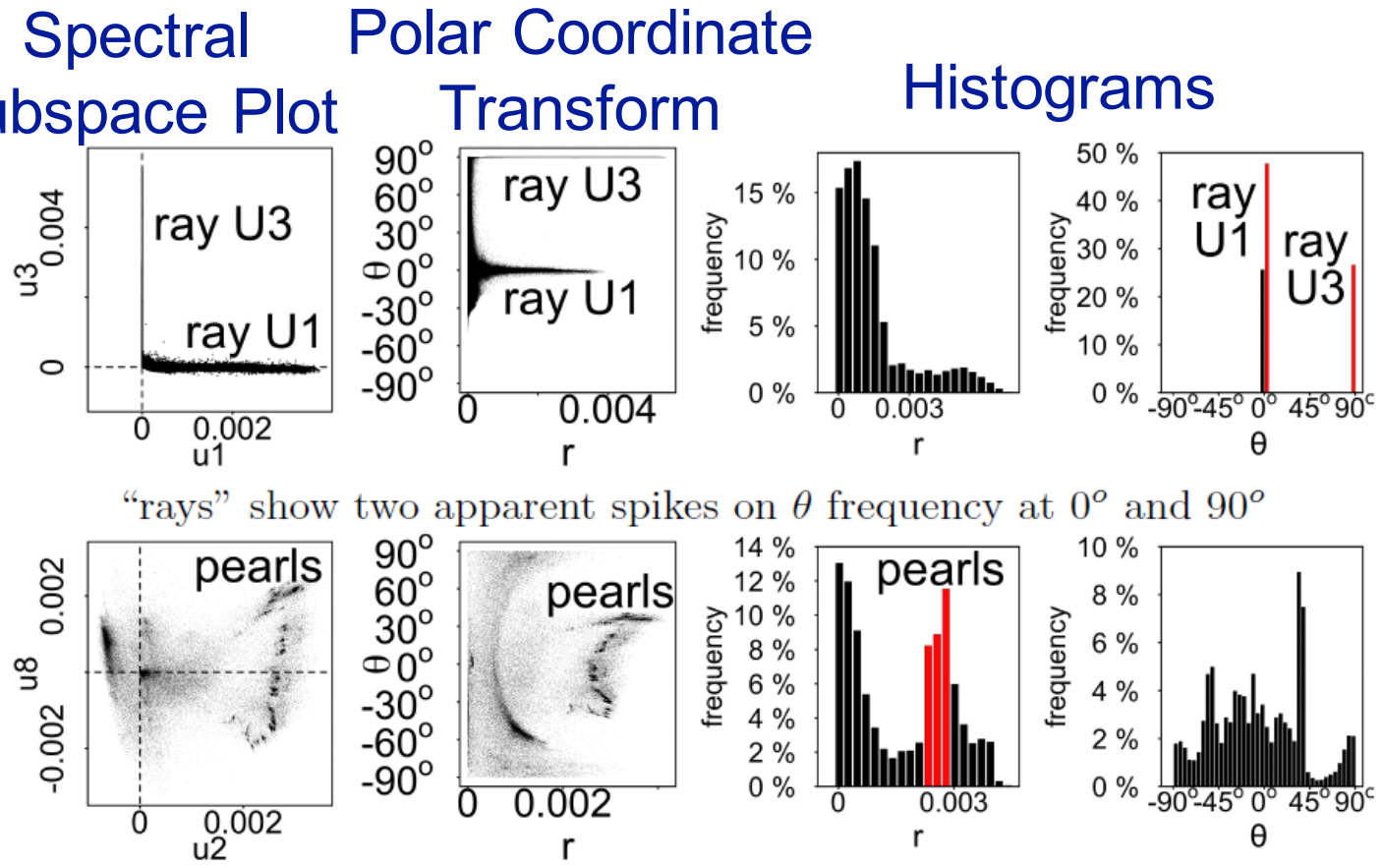
Spectral Subspace Plot: Case #4

- Overlapping: "Staircase", "Pearls"



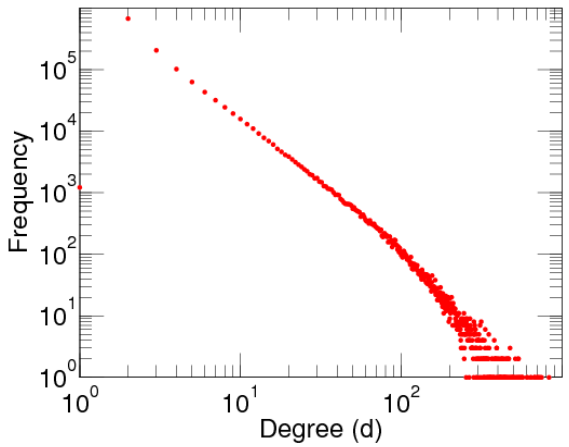
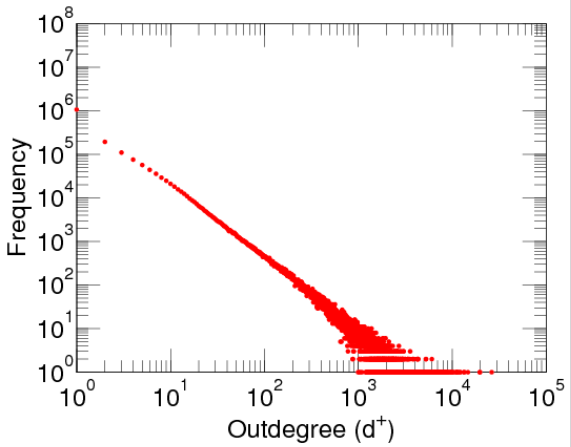
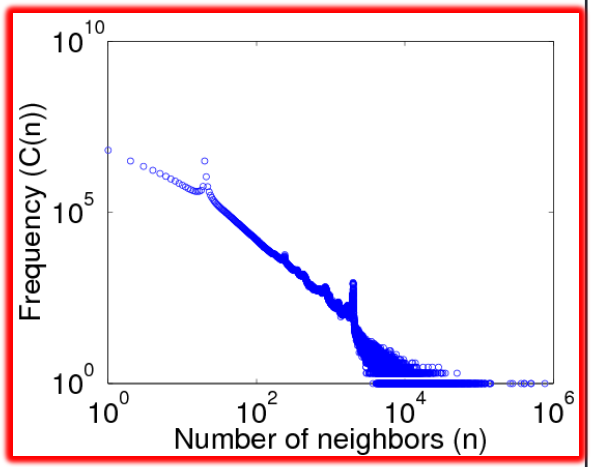
Rule 4 ("pearls"): a "staircase" of three partially overlapping blocks.

Spectral Subspace Plot: Reading & LockInfer

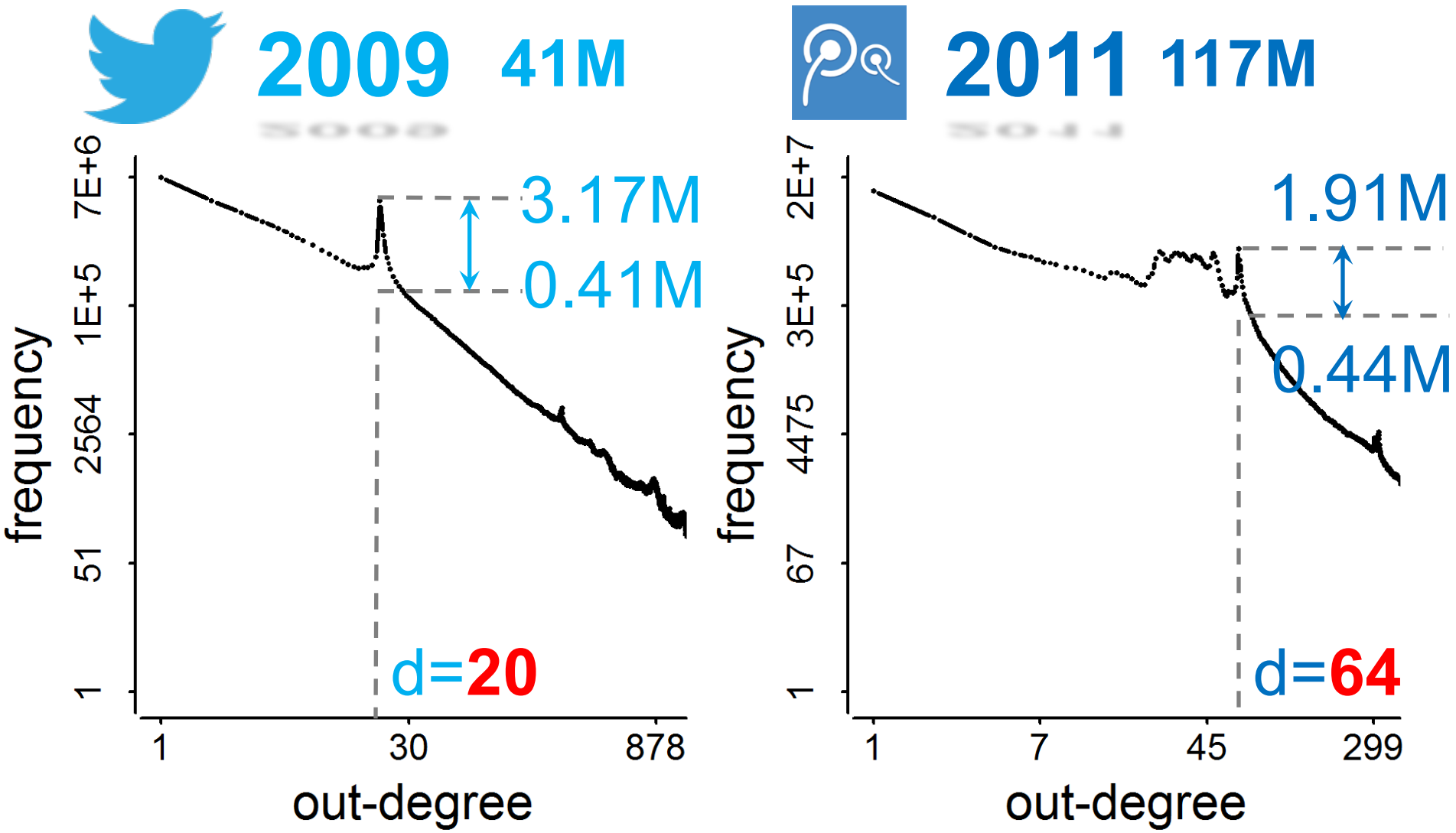


Graphical View: Power-Law Distribution

- Out-degree distribution

DBLP **Author**-publication	**Flickr** **User**-user	**Twitter** **Who**-follows-whom
		

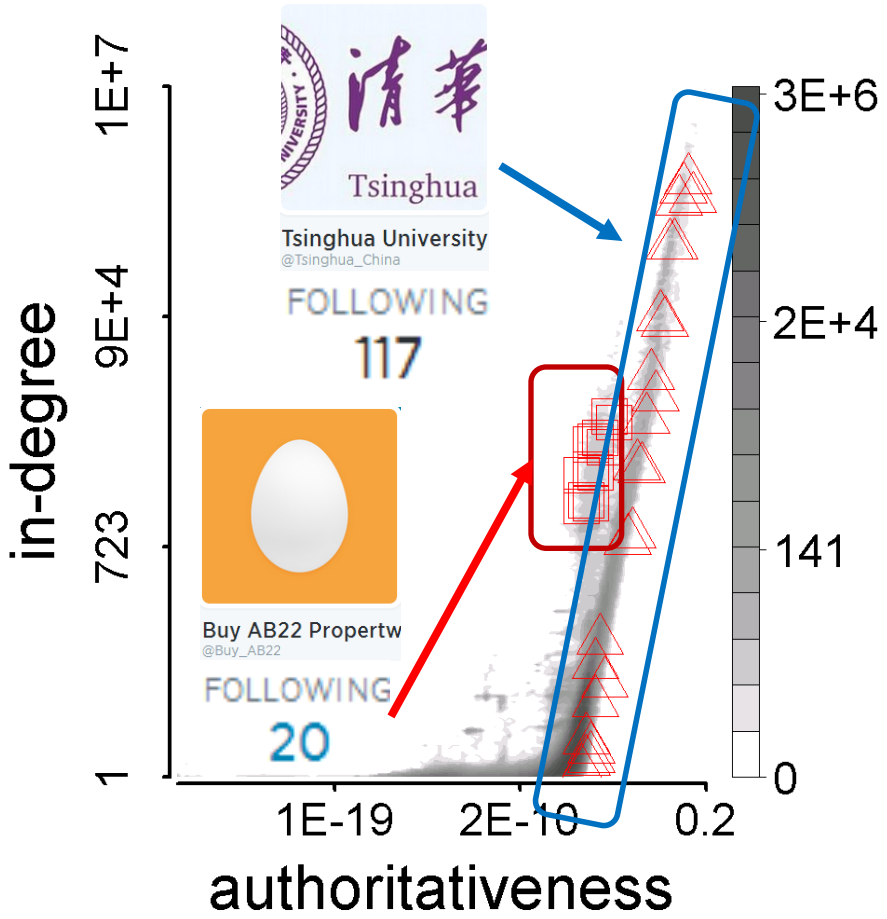
[konect.uni-koblenz.de/networks/]



Jiang et al. **CatchSync: Catching Synchronized Behavior in Large Directed Graphs.** *KDD Best Paper Finalist*, 2014.

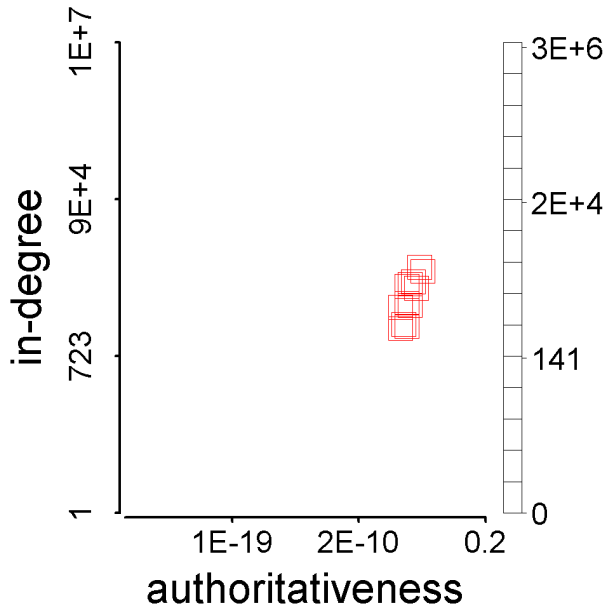
Synchronized Behavior: Features

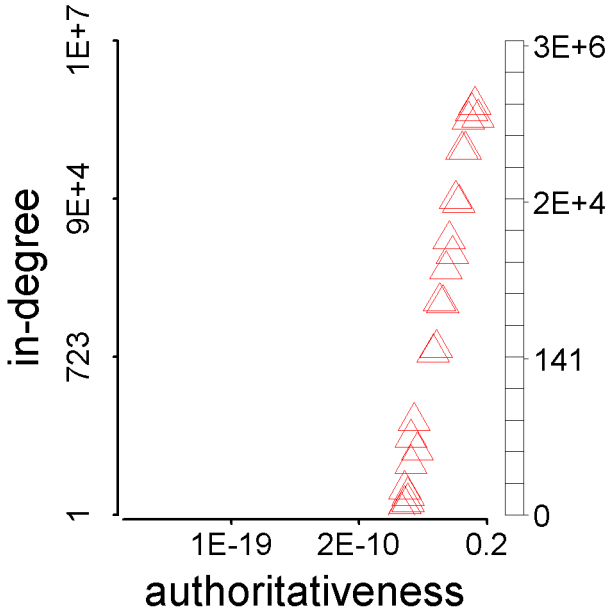
- Synchronized
- Abnormal



Synchronized Behavior: Synchronicity

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

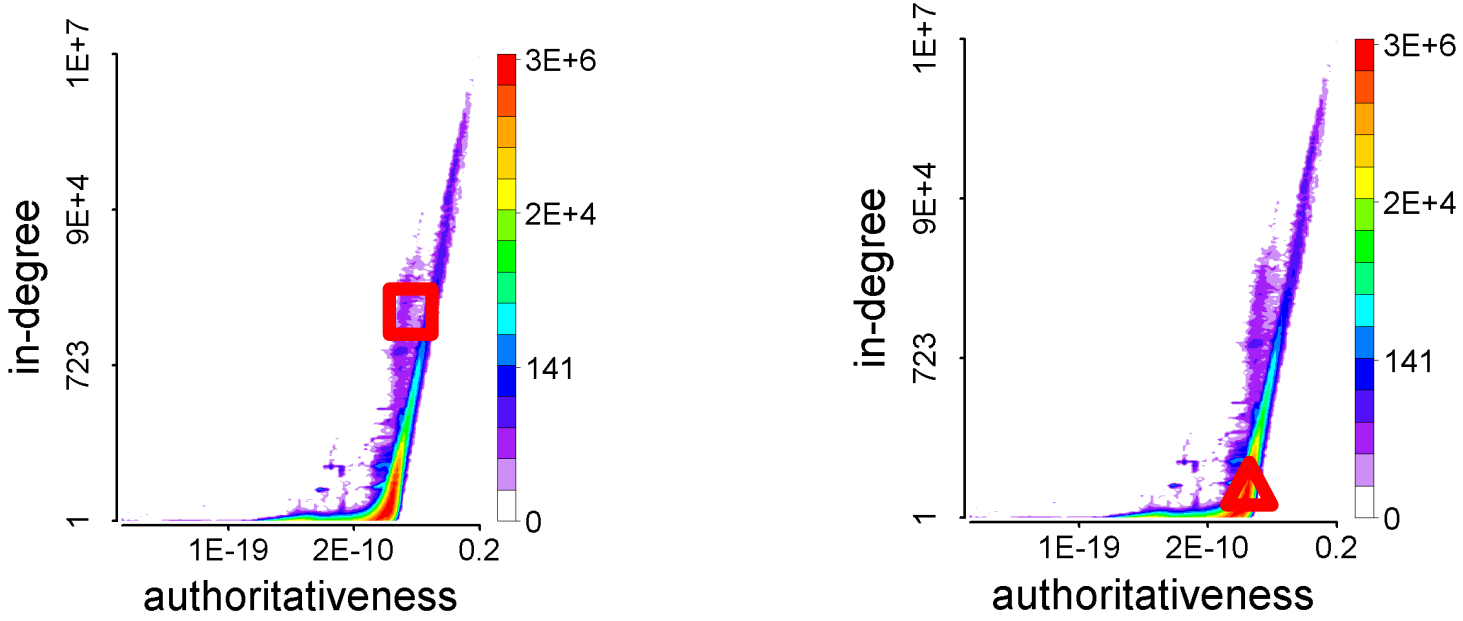




Jiang et al. **CatchSync: Catching Synchronized Behavior in Large Directed Graphs.** *KDD Best Paper Finalist*, 2014.

Synchronized Behavior: Normality

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

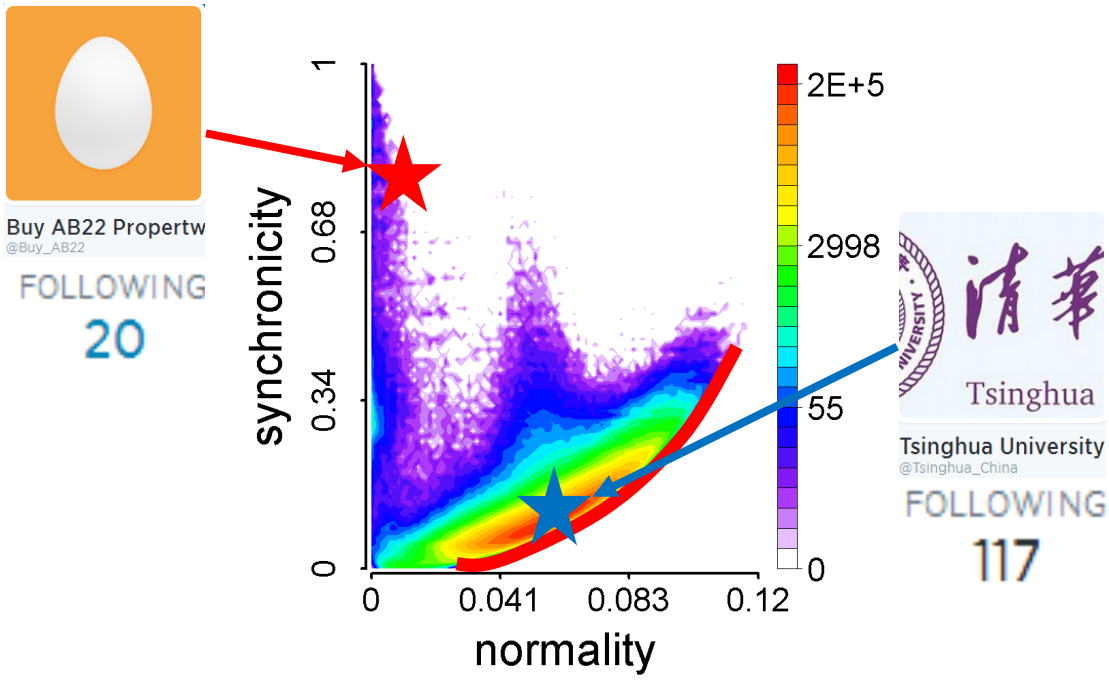


Jiang et al. **CatchSync: Catching Synchronized Behavior in Large Directed Graphs.** *KDD Best Paper Finalist*, 2014.

Synchronized Behavior: Theorem & CatchSync

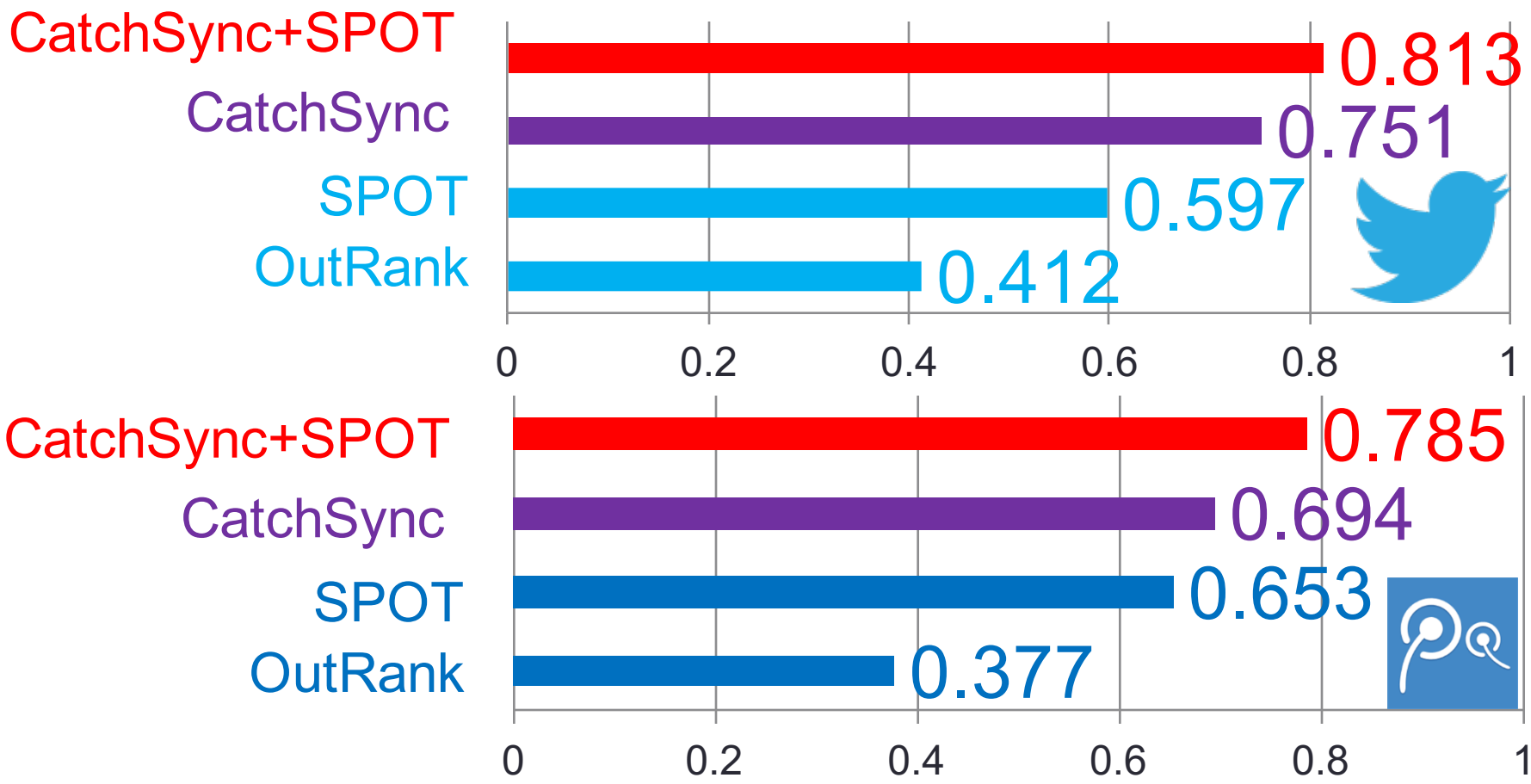
❖Synchronicity-Normality Plot

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



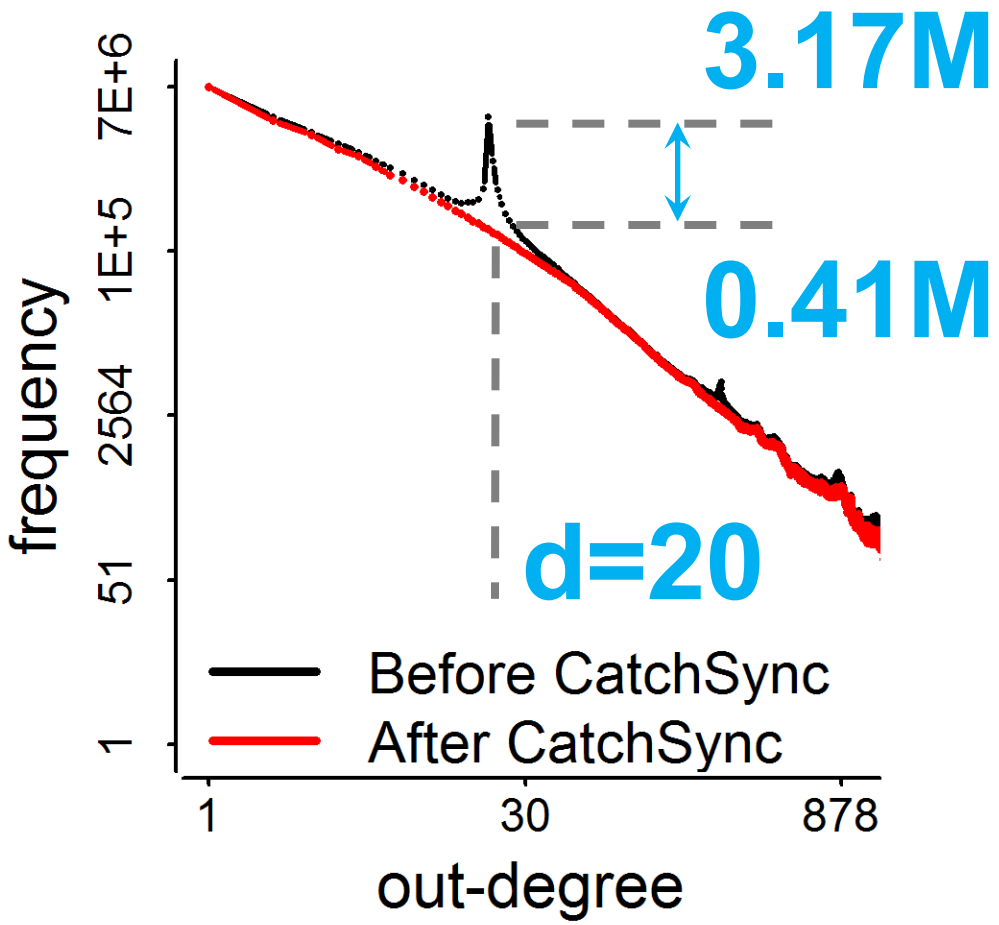
Jiang et al. **CatchSync: Catching Synchronized Behavior in Large Directed Graphs.** *KDD Best Paper Finalist*, 2014.

Synchronized Behavior: Performance



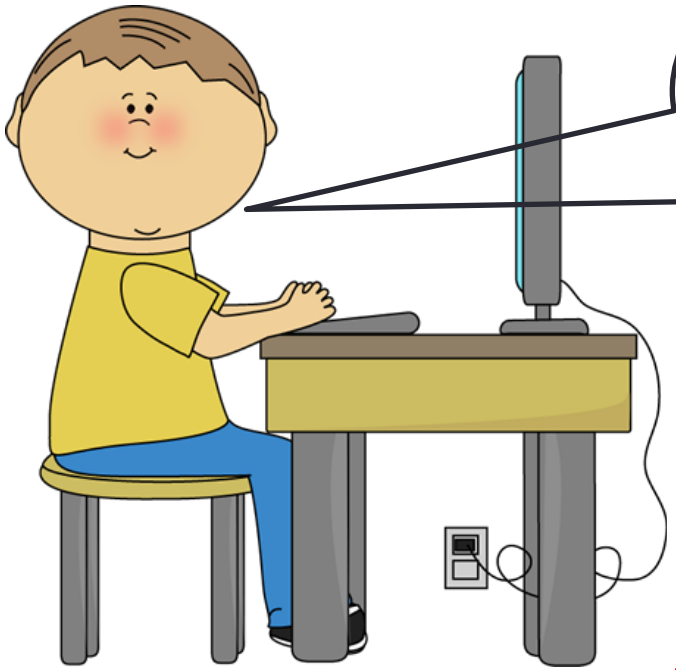
Jiang et al. **CatchSync: Catching Synchronized Behavior in Large Directed Graphs.** *KDD Best Paper Finalist*, 2014.

Synchronized Behavior: Performance



Jiang et al. **CatchSync: Catching Synchronized Behavior in Large Directed Graphs.** *KDD Best Paper Finalist*, 2014.

Beyond Graph: Multi-Dimensional Fraud



My boss wants me to **catch fraud** in such a big table – **billions of records, tens of columns**!!! How?!

	ID	USER_NAME	CREATED_AT	TEXT	HASH_TAGS
1	251	SpiritSofts	Dec 14, 2013	SAP HANA ONLINE TRAINING COURSE CONTENT http://t.co/2DefOMC0Vi	
2	252	Blue net studiO	Dec 14, 2013	sap hana online training and placenet 2 http://t.co/S1wGh8n5Kk	
3	253	Hana Kingham	Dec 14, 2013	Right film fest today: love actually, elf, gravity, training day. #dayyym	dayyym,
4	254	Nora Apnila J...	Dec 14, 2013	Alhamdulilaaahhhh...selesai ikutin kelanjutan training dadakan mb Hana ...	
5	255	ZaranTech	Dec 14, 2013	I added a video to a @YouTube playlist http://t.co/O3qD9wfi8K SAP BUSI...	
6	256	ZaranTech	Dec 14, 2013	I added a video to a @YouTube playlist http://t.co/XxrfuCUqAS SAP BUSI...	
7	257	Helmich op t...	Dec 14, 2013	Reserveer alvast 15 januari 2014 training HANA Essentials #SAP #HANA	SAP,HANA,
8	258	Social News	Dec 13, 2013	sap hana online training and placenet 2 http://t.co/JlaA41IdnV	
9	259	Nurjanah	Dec 13, 2013	Baca notif fb .. ada training dadakaann dari eyang kita.... ayo wara wiri ca...	
10	260	Nora Apnila J...	Dec 13, 2013	lanjutt di rumaah dulu ikutan trainingnyaaa..mau buru buru pulang see u...	
11	261	madhu	Dec 13, 2013	SAP HANA TRAINING | SAP HANA PLACEMENT | SAP HANA INSTITUTE I...	
12	262	Hana O'Neill	Dec 13, 2013	@sarahsilvanator no I have life guard training Saturday and my final test t...	
13	263	arjun	Dec 13, 2013	sap grc online training |sap hana sap security online training@YEKTEK - A...	

fraud

Jiang et al. **A General Suspicious Metric for Dense Blocks in Multi-Modal Data.** *ICDM*, 2015.

Multi-Dimensional Fraud

Wholesalebargain2015 Retweeted

Real Time Deals @ebayrt · 2h

Seattle Mariners Mlb #Majestic Authentic Diamond Blue Stitched 2 Xl M... (Sanford) USD 25 ebayrt.co/sports-mem-car… #eBay #eBayUS via @wil30225

2

Wholesalebargain2015 Retweeted

Real Time Deals @ebayrt · 2h

Embroidered Navy Blue Aztec Mexican Top/ Long Sleeve Ethnic Mod... USD 35 ebayrt.co/clothing-shoes… #Handmade #eBay #eBayUS via @smilingbluedog

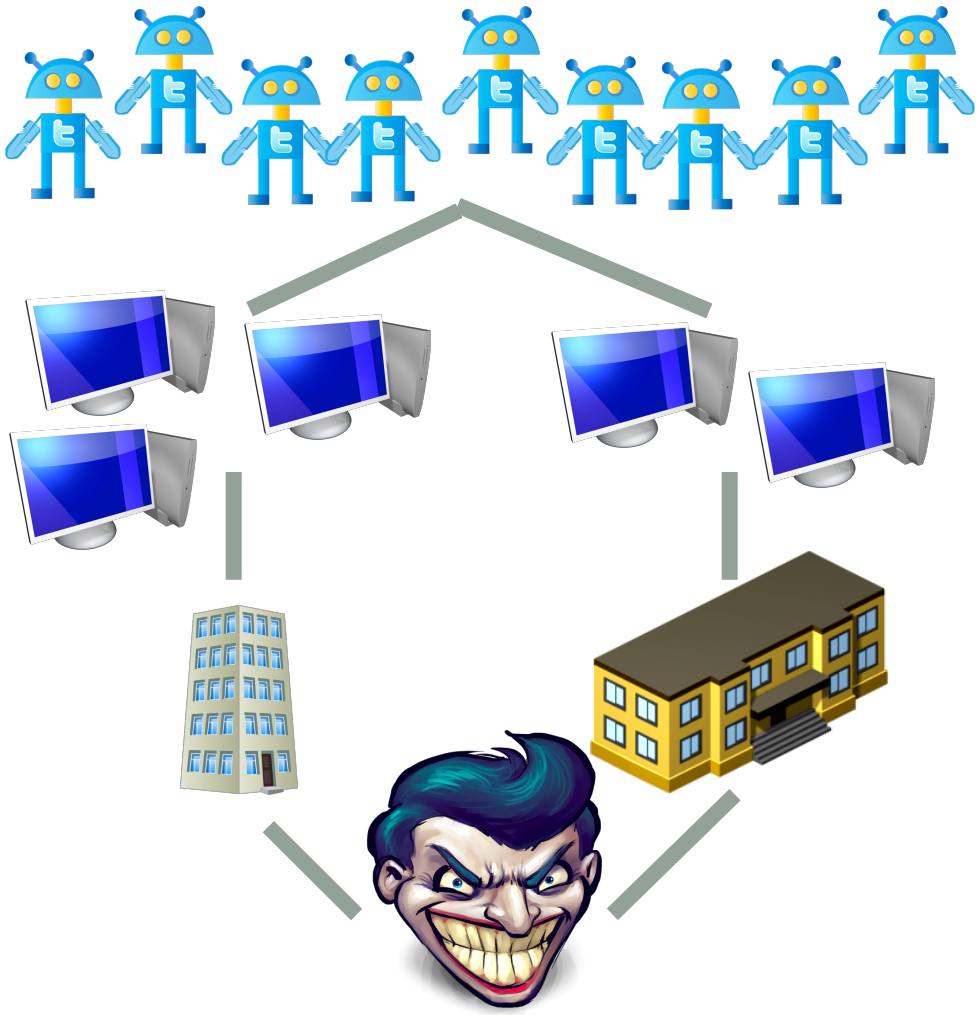
4

Wholesalebargain2015 Retweeted

Real Time Deals @ebayrt · 1h

Contractubex Children Cartoon Boxing Gloves Red (Bloomington) USD 21.78 ebayrt.co/sporting-goods… #eBay #eBayUS via @GaroldFrenz

4



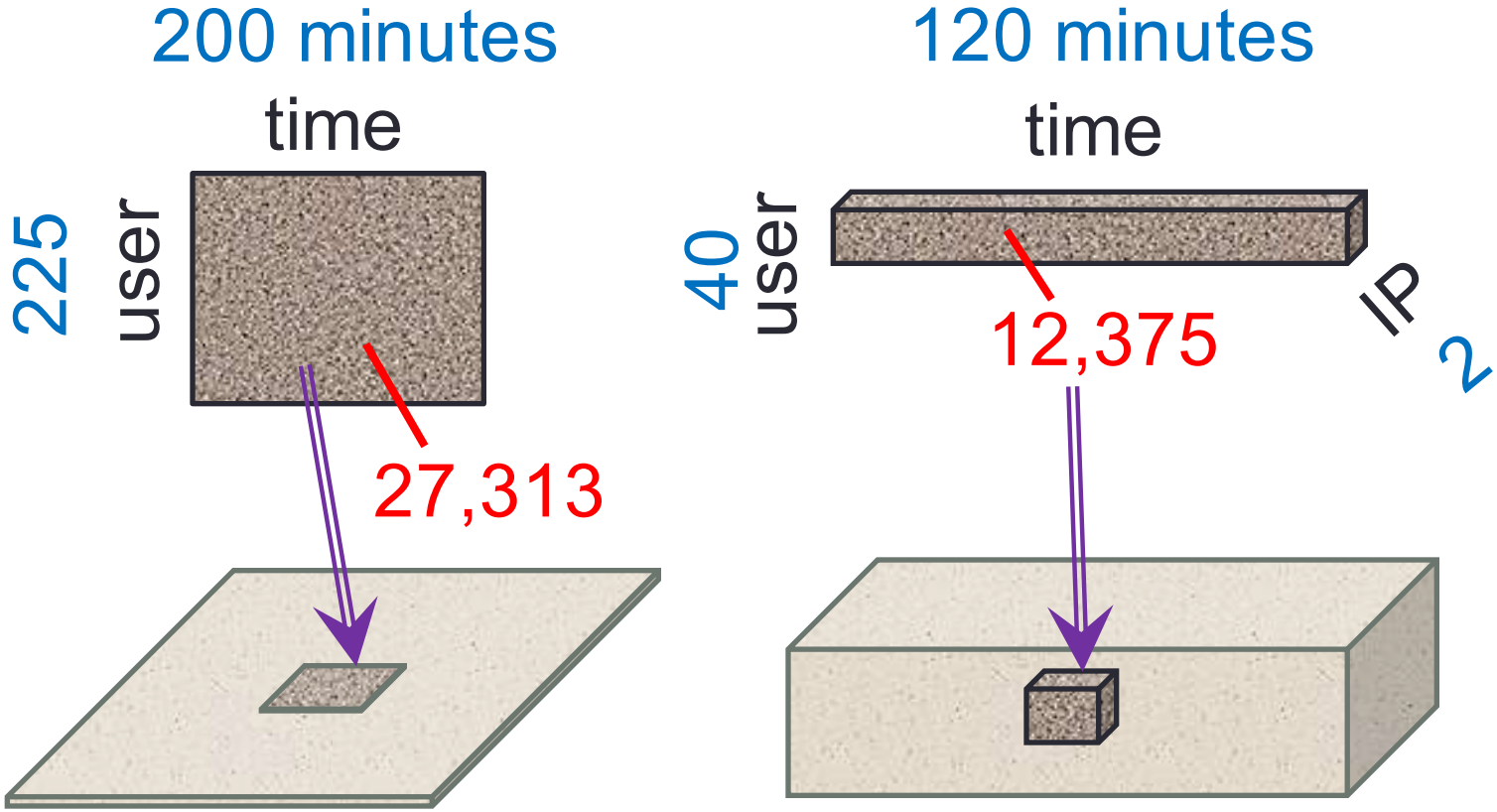
Jiang et al. **A General Suspicious Metric for Dense Blocks in Multi-Modal Data.** *ICDM*, 2015.

Multi-Dimensional Fraud

Dataset	Mode				Mass
Retweeting	User	Root ID	IP	Time (min)	#retweet
	29.5M	19.8M	27.8M	56.9K	211.7M
Trending (Hashtag)	User	Hashtag	IP	Time (min)	#tweet
	81.2M	1.6M	47.7M	56.9K	276.9M
Network attacks (LBNL)	Src-IP	Dest-IP	Port	Time (sec)	#packet
	2,345	2,355	6,055	3,610	230,836

Jiang et al. **A General Suspicious Metric for Dense Blocks in Multi-Modal Data.** *ICDM*, 2015.

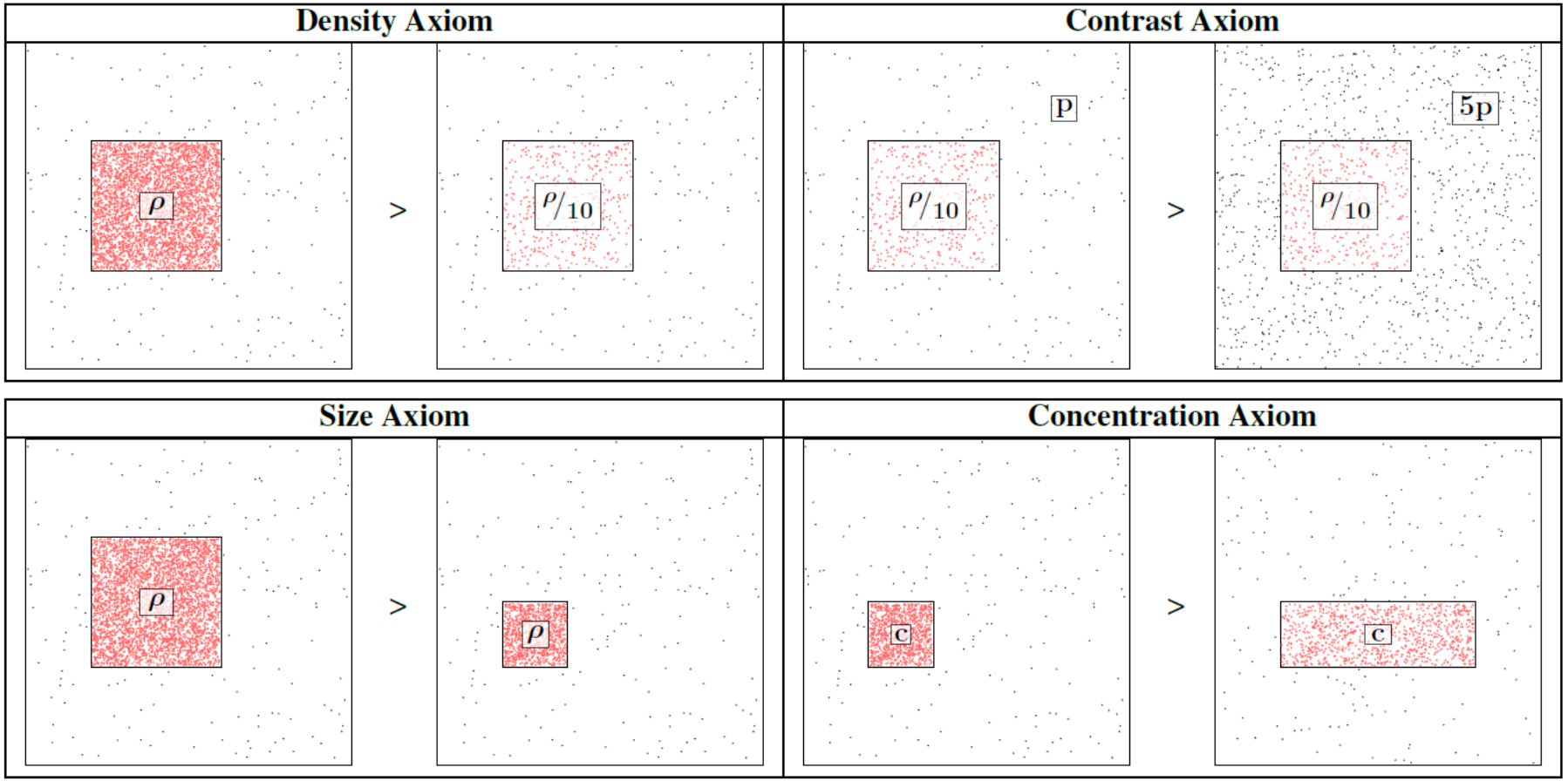
Suspiciousness: Density in Multi-Dimensions



Question: Which is more suspicious?

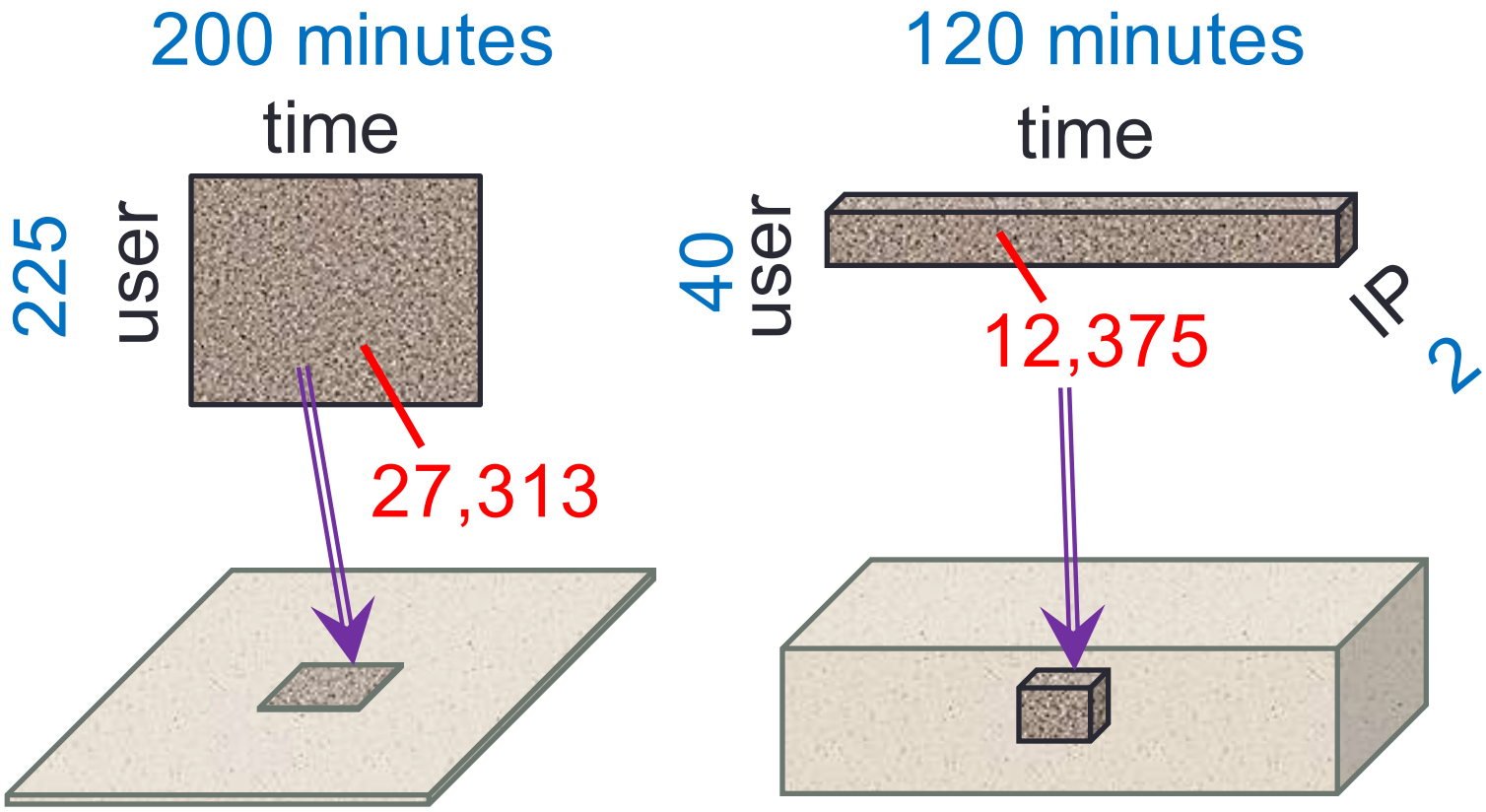
- Before we search, we should be able to rank.

Suspiciousness: Axioms



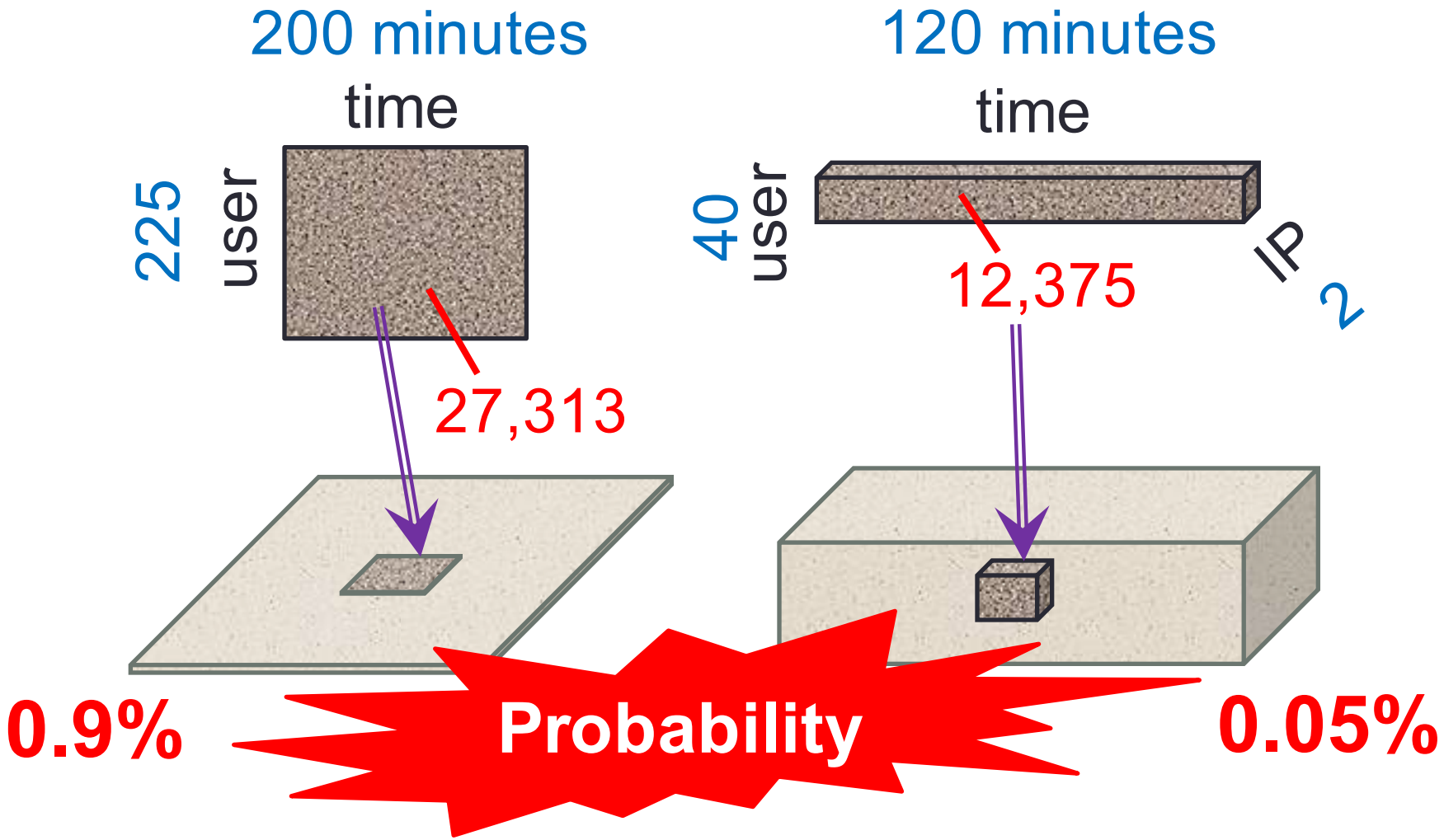
Jiang et al. **A General Suspicious Metric for Dense Blocks in Multi-Modal Data.** *ICDM*, 2015.

Suspiciousness: General Metric & CrossSpot



Jiang et al. **A General Suspicious Metric for Dense Blocks in Multi-Modal Data.** *ICDM*, 2015.

Suspiciousness: General Metric & CrossSpot



Suspiciousness: General Metric & CrossSpot

- Negative log likelihood of block's probability

$$f(n, c, N, C) = -\log\left[Pr(Y_n = c)\right]$$

Lemma *Given an $n_1 \times \cdots \times n_K$ block of mass c in $N_1 \times \cdots \times N_K$ data of total mass C, the suspiciousness function is*

$$f(\mathbf{n}, c, \mathbf{N}, C) = c(\log \frac{c}{C} - 1) + C \prod_{i=1}^{K} \frac{n_i}{N_i} - c \sum_{i=1}^{K} \log \frac{n_i}{N_i}$$

Using ρ as the block's density and p is the data's density, we have the simpler formulation

$$\hat{f}(\mathbf{n}, \rho, \mathbf{N}, p) = \left(\prod_{i=1}^{K} n_i\right) D_{KL}(\rho || p)$$

Jiang et al. **A General Suspicious Metric for Dense Blocks in Multi-Modal Data.** *ICDM*, 2015.

Suspiciousness: Trend Manipulating

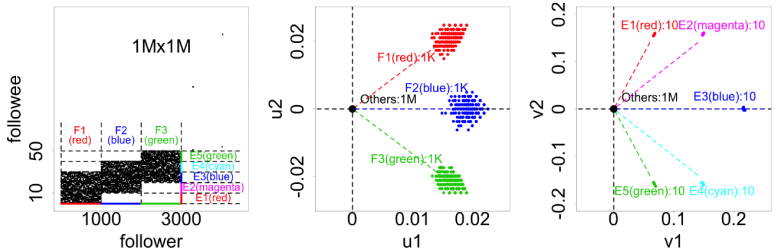
User × hashtag × IP × minute	Mass c	Suspiciousness
582 × 3 × 294 × **56,940**	5,941,821	111,799,948
188 × 1 × 313 × **56,943**	2,344,614	47,013,868
75 × 1 × 2 × 2,061	689,179	19,378,403

User ID	Time	IP address (city, province)	Tweet text with hashtag
USER-D	11-18 12:12:51	IP-1 (Deyang, Shandong)	**#Snow#** the Samsung GALAXY SII QQ Service customized version...
USER-E	11-18 12:12:53	IP-1 (Deyang, Shandong)	**#Snow#** the Samsung GALAXY SII QQ Service customized version...
USER-F	11-18 12:12:54	IP-2 (Zaozhuang, Shandong)	**#Snow#** the Samsung GALAXY SII QQ Service customized version...
USER-E	11-18 12:17:55	IP-1 (Deyang, Shandong)	**#Li Ning - a weapon with a hero#** good support activities!
USER-F	11-18 12:17:56	IP-2 (Zaozhuang, Shandong)	**#Li Ning - a weapon with a hero#** good support activities!
USER-D	11-18 12:18:40	IP-1 (Deyang, Shandong)	**#Toshiba Bright Daren#** color personality test to find out your sense...
USER-E	11-18 17:00:31	IP-2 (Zaozhuang, Shandong)	**#Snow#** the Samsung GALAXY SII QQ Service customized version...
USER-D	11-18 17:00:49	IP-2 (Zaozhuang, Shandong)	**#Toshiba Bright Daren#** color personality test to find out your sense...
USER-F	11-18 17:00:56	IP-2 (Zaozhuang, Shandong)	**#Li Ning - a weapon with a hero#** good support activities!

Jiang et al. **A General Suspicious Metric for Dense Blocks in Multi-Modal Data.** *ICDM*, 2015.

Summary for Detecting Unnatural Behavior

- Suspicious behavioral patterns
  - **Lockstep pattern:** CopyCatch, LockInfer
  - **Synchronized pattern:** CatchSync
  - **Multi-dimensional suspiciousness:** CrossSpot



Rule 4 ("pearls"): a "staircase" of three partially overlapping blocks.