

BEHAVIORAL MODELING IN SOCIAL NETWORKS FROM MICRO TO MACRO

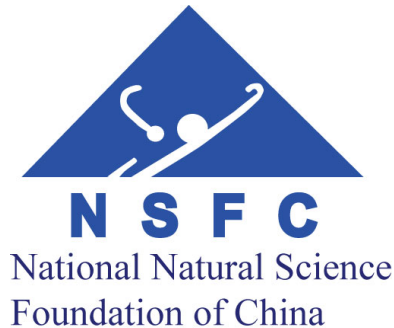
Meng Jiang, *University of Illinois at Urbana-Champaign*

Peng Cui, *Tsinghua University*

ICDM 2015 TUTORIAL
Atlantic City, NJ



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,

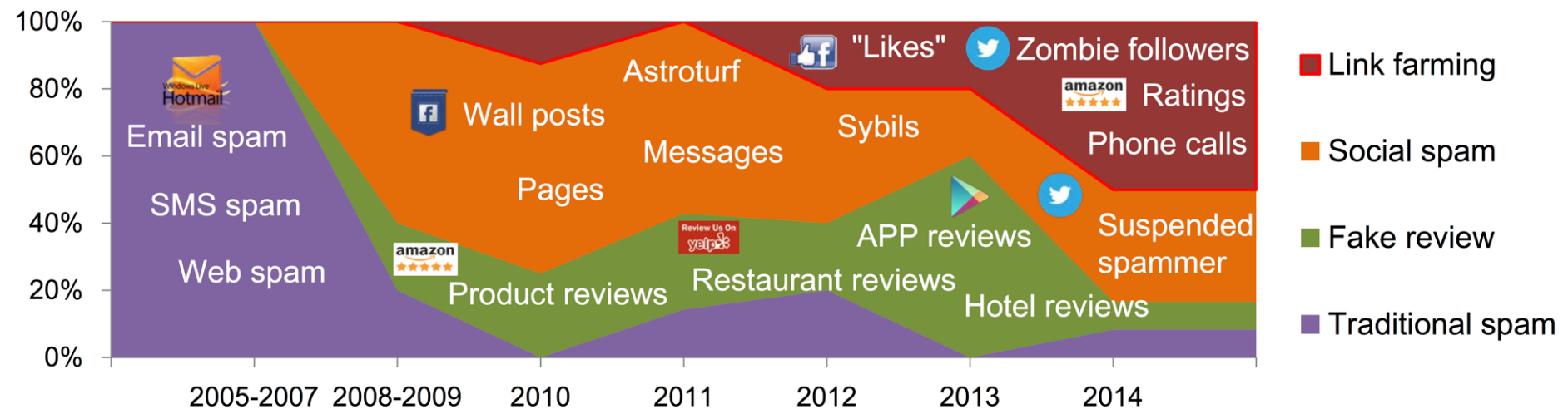
WBS:R-252-300-001-490

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)

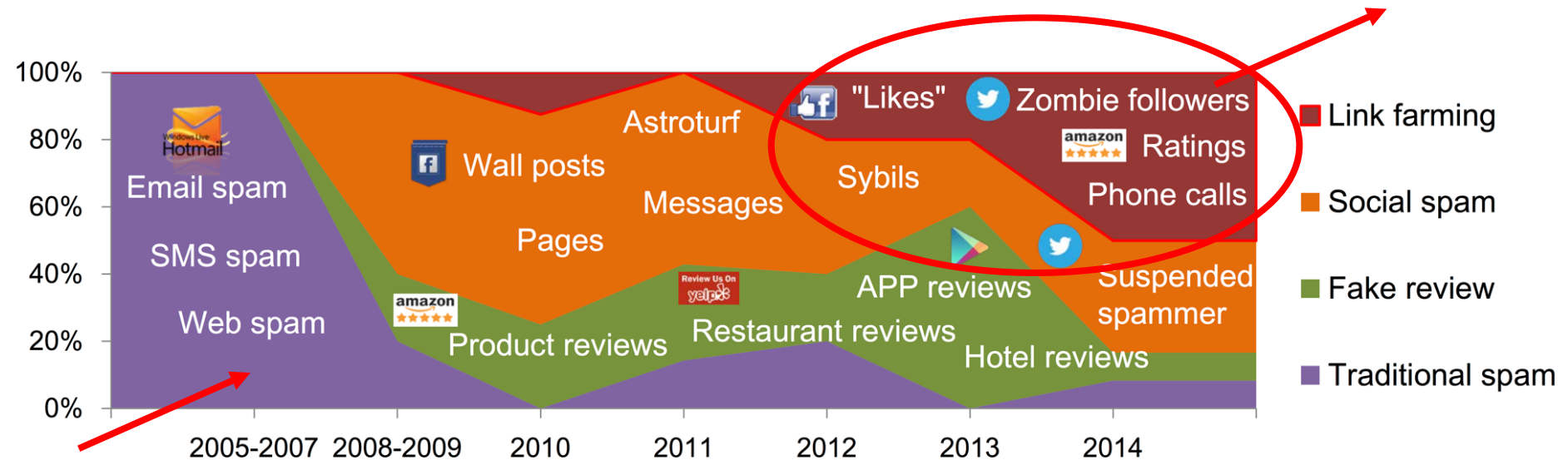


FALOUTSOS,
Christos
(CMU)

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

Social Link Farming



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)

Social Link Farming

❖ Selling Twitter followers

Followers	Price	Delivery	Save +
5,000 FOLLOWERS 400 FREE	\$69.99	Delivery within 3-4 days	3%
2,000 FOLLOWERS 300 FREE	\$29.99	Delivery within 2-3 days	2%
1,000 FOLLOWERS 200 FREE	\$15.99	Delivery within 1-2 days	-
10,000 FOLLOWERS 500 FREE	\$119.99	Delivery within 4-5 days	14%
20,000 FOLLOWERS 1000 FREE	\$229.99	Delivery within 5-8 days	34%

Social Link Farming

❖ Selling Facebook Likes

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

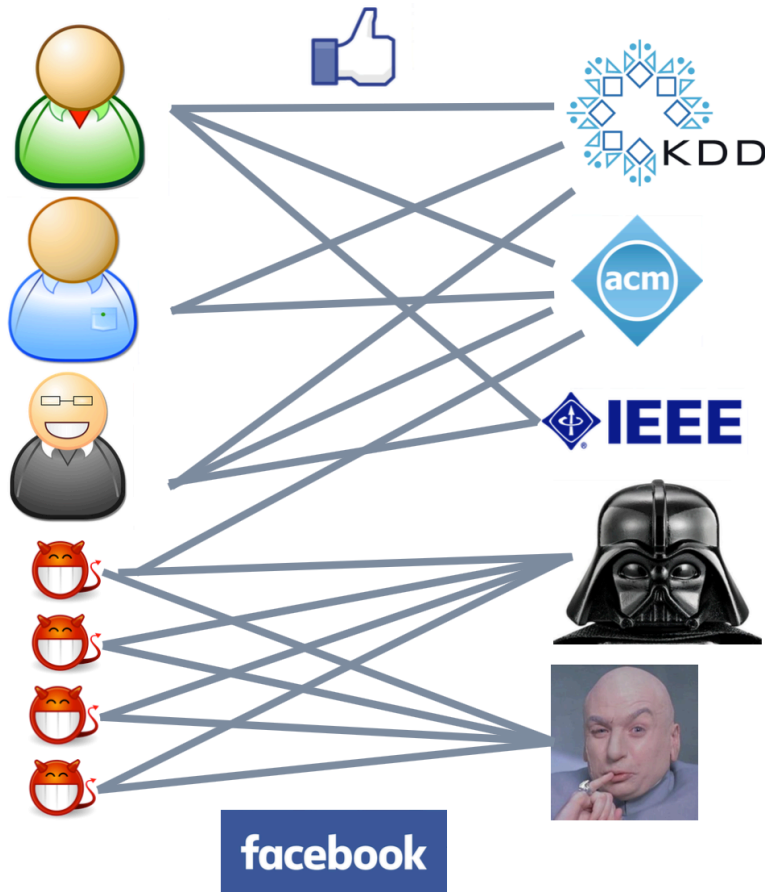
Suspicious Behavior Detection

- ❖ 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

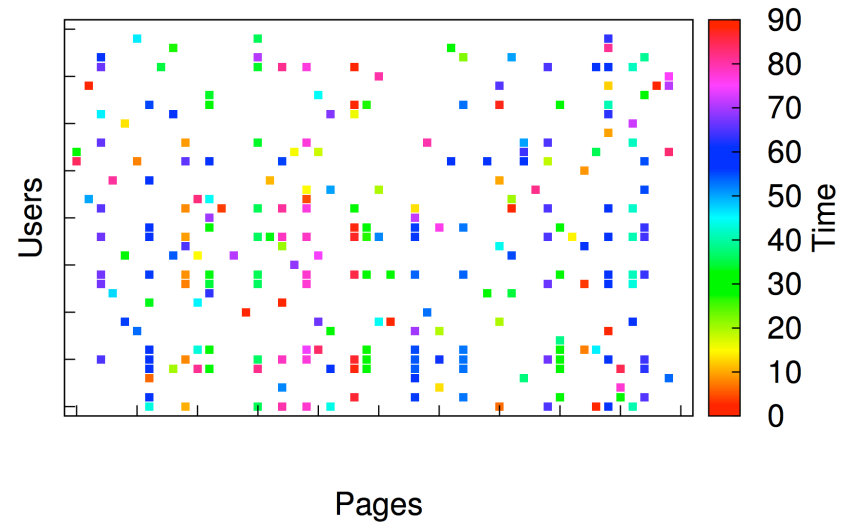
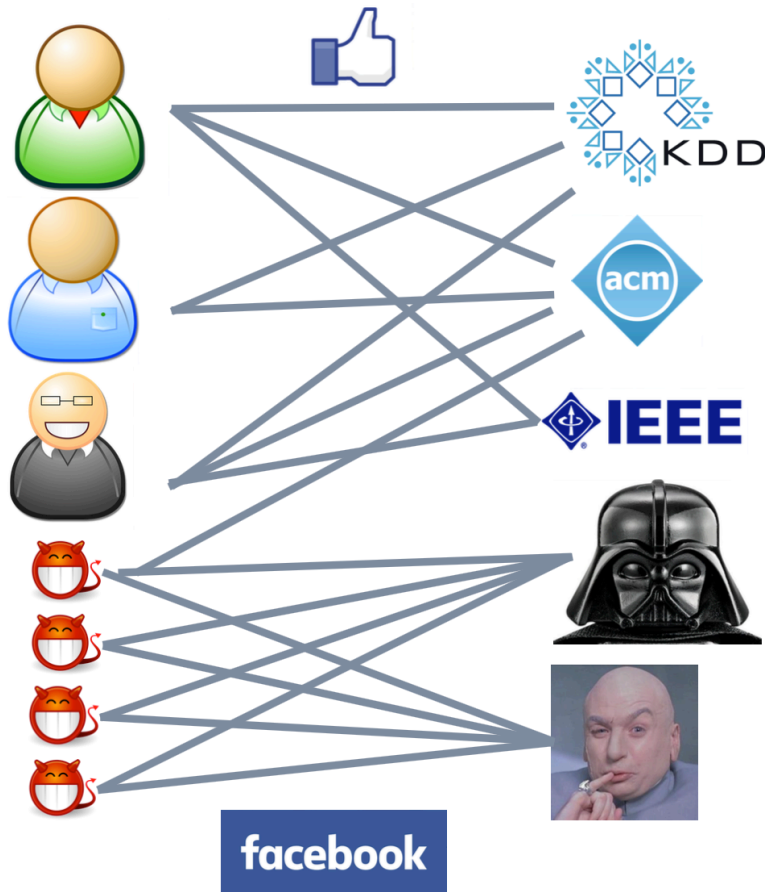
Lockstep Behavior: Facebook Likes

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

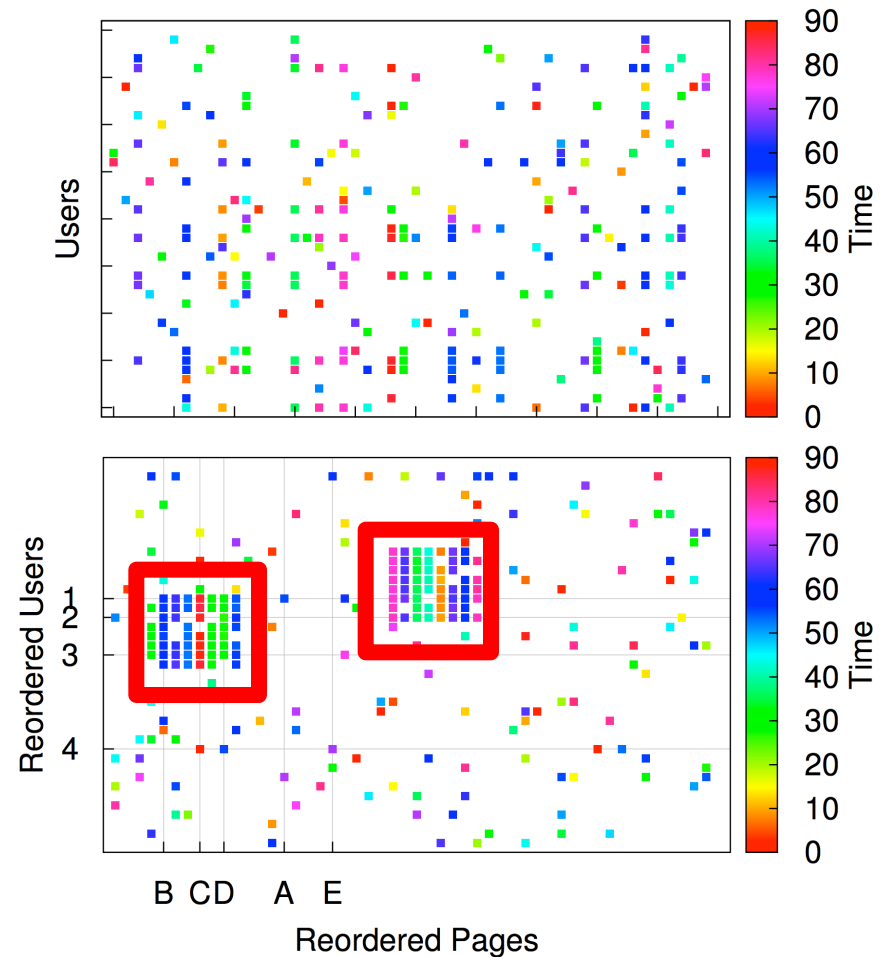
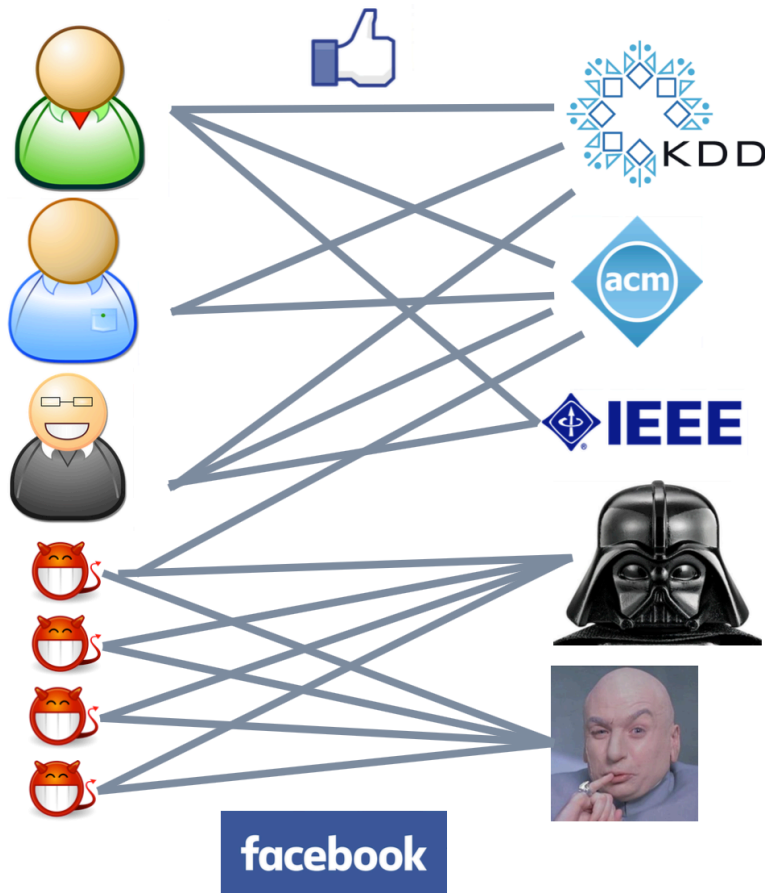
Lockstep Behavior: Facebook Likes



Lockstep Behavior: Graphical View



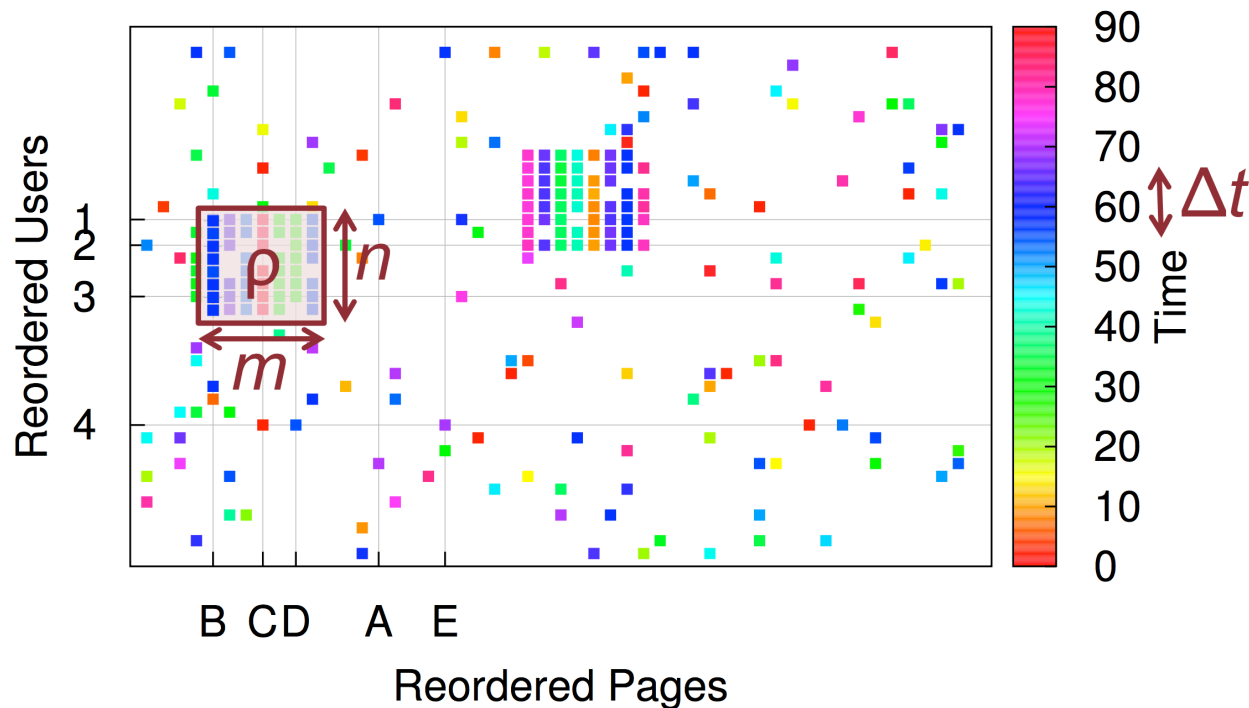
Lockstep Behavior: Reorder Matrix



Lockstep Behavior: Seed + Search

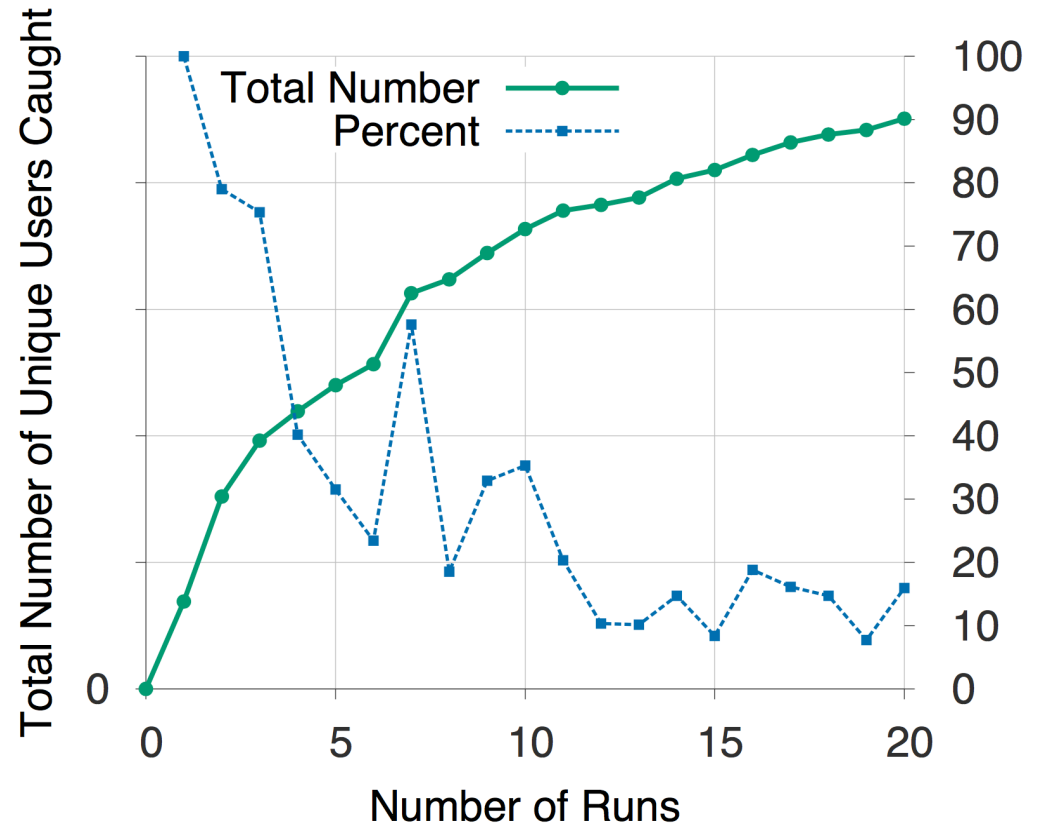
❖ CopyCatch

❖ “Near Bipartite Core”: n users, m Pages, ρ , Δt



Lockstep Behavior: Performance

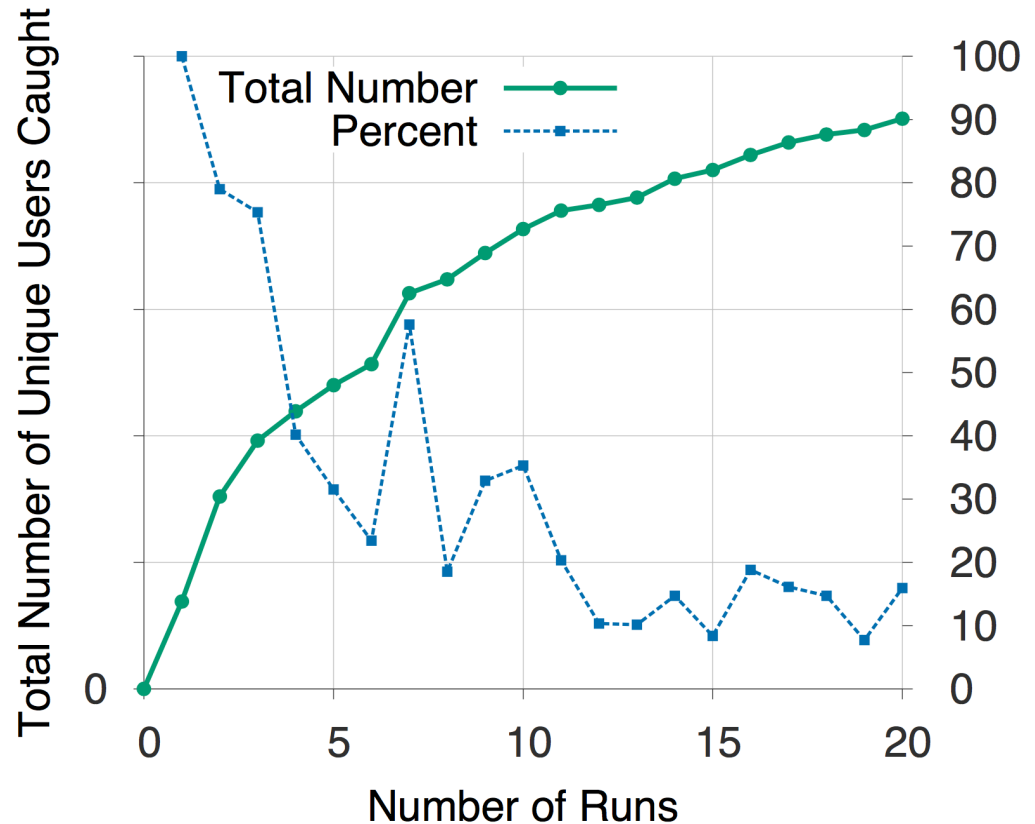
❖ CopyCatch



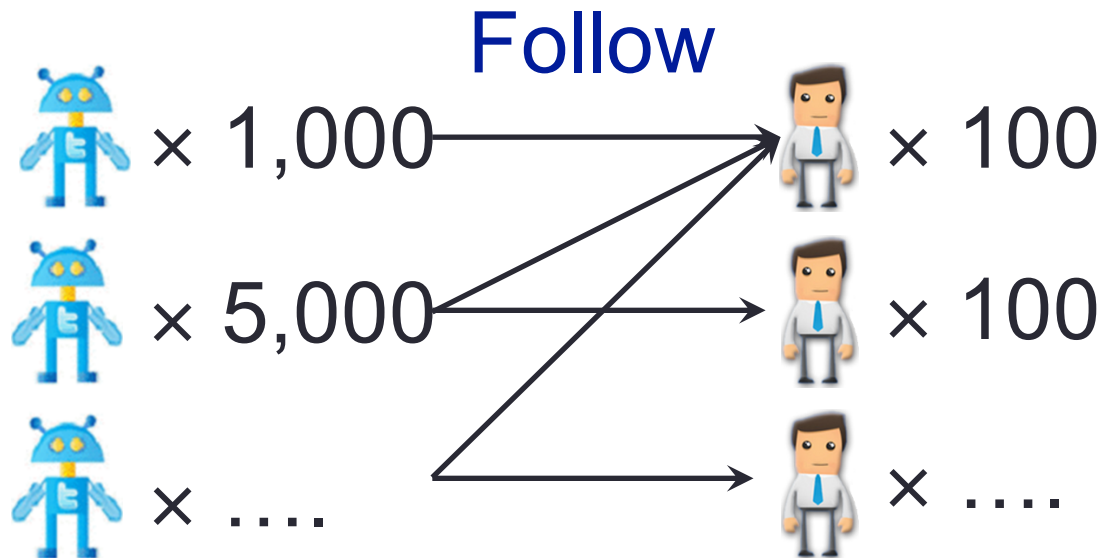
Lockstep Behavior: Performance

❖ CopyCatch

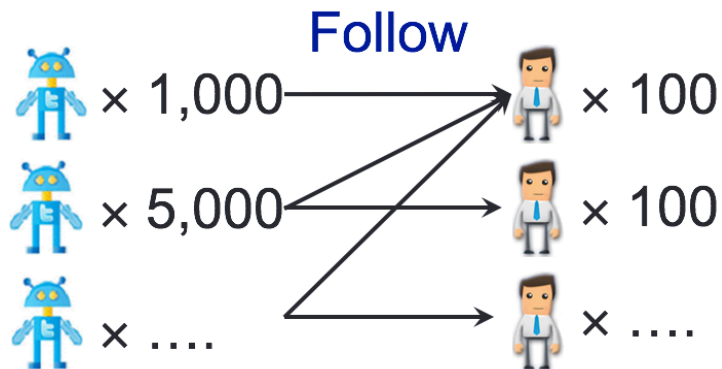
Seed Selection!!!



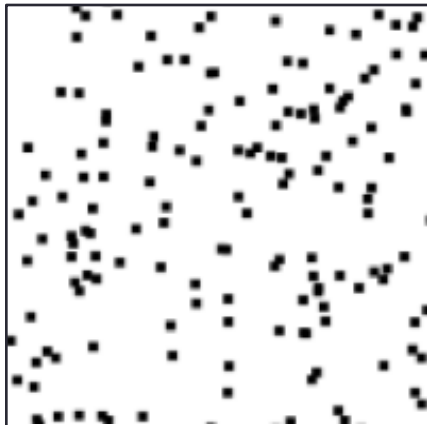
Lockstep Behavior: **Twitter Followers**



Lockstep Behavior: Reorder Matrix

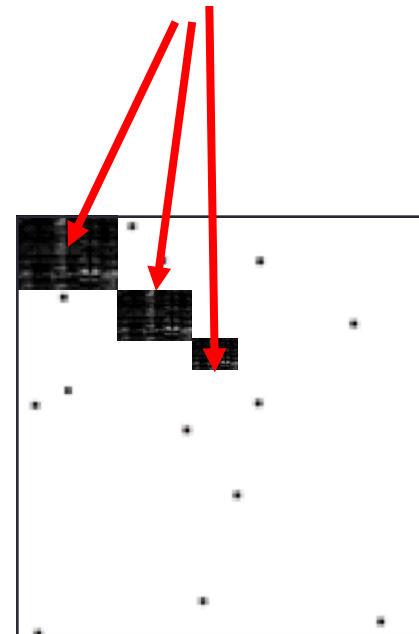


Given

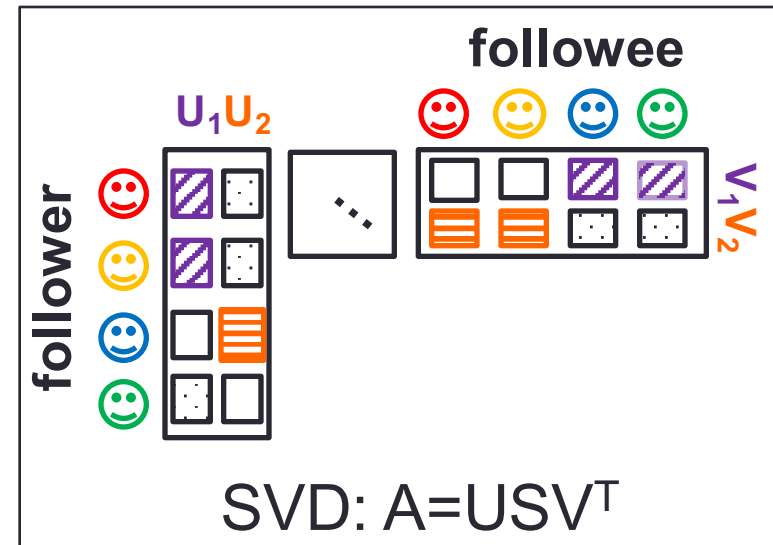
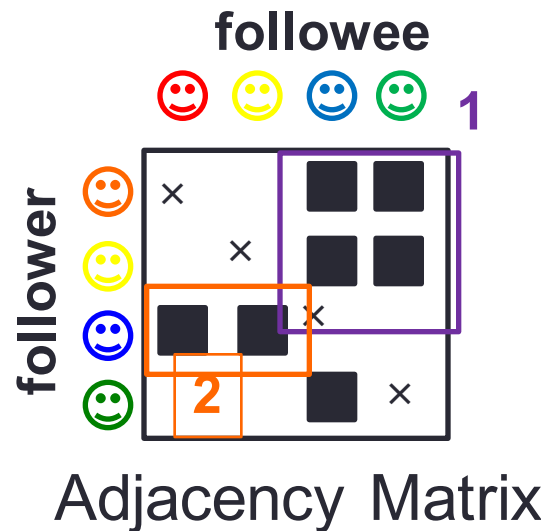
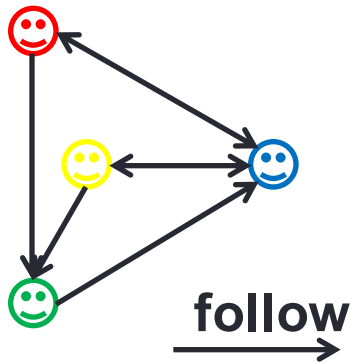


Reorder

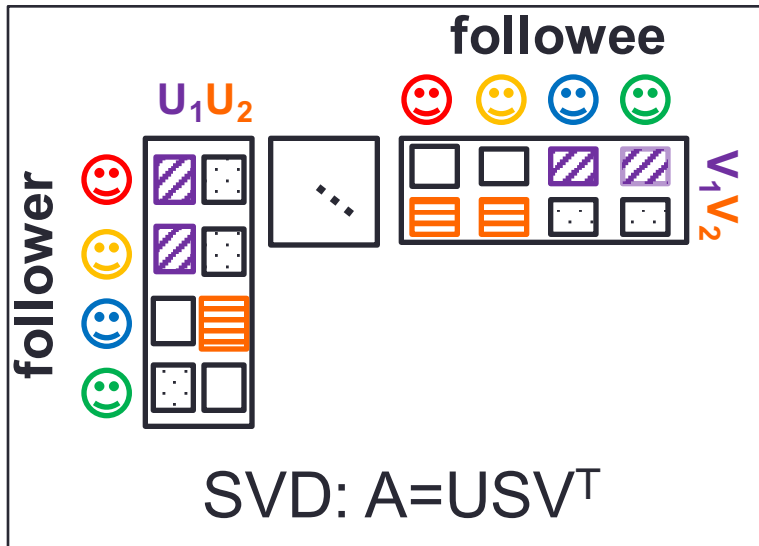
Seeds



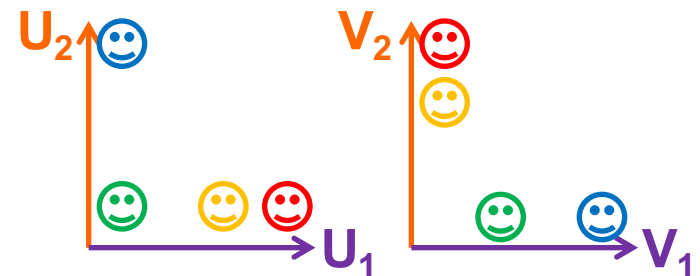
Lockstep Behavior: SVD Reminder



Lockstep Behavior: Spectral Subspace



Pairs of singular vectors:

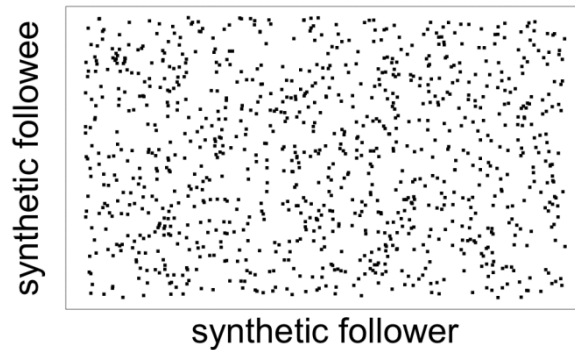


“Spectral Subspace Plot”

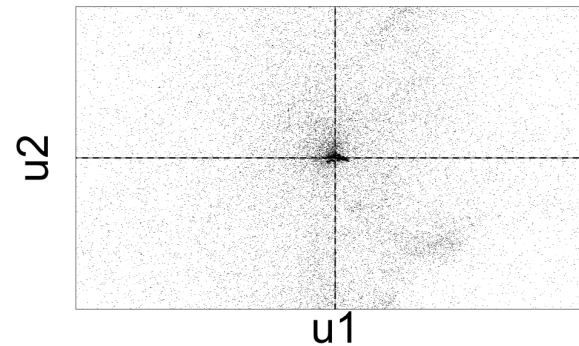
Spectral Subspace Plot: Case #0

❖ NO lockstep behavior: Scatter

Adjacency Matrix



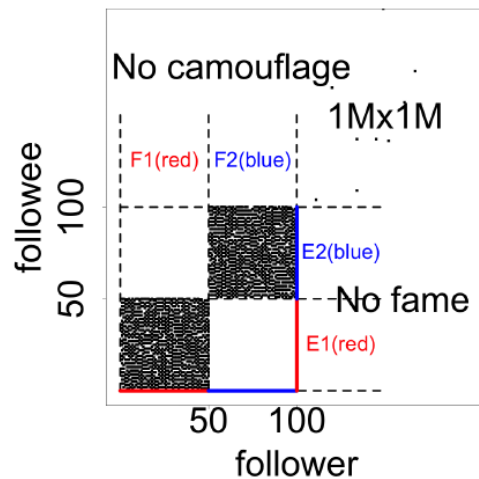
Spectral Subspace Plot



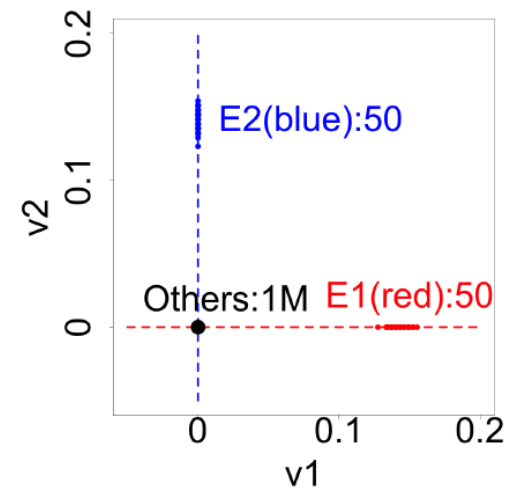
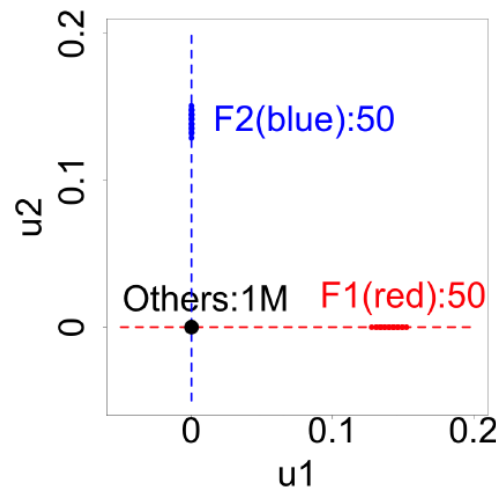
Spectral Subspace Plot: Case #1

❖ Non-overlapping lockstep: “Rays”

Adjacency Matrix



Spectral Subspace Plot

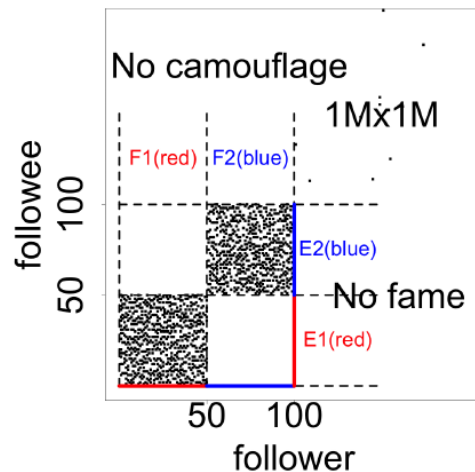


Rule 1 (short “rays”): two blocks, high density (90%), no “camouflage”, no “fame”

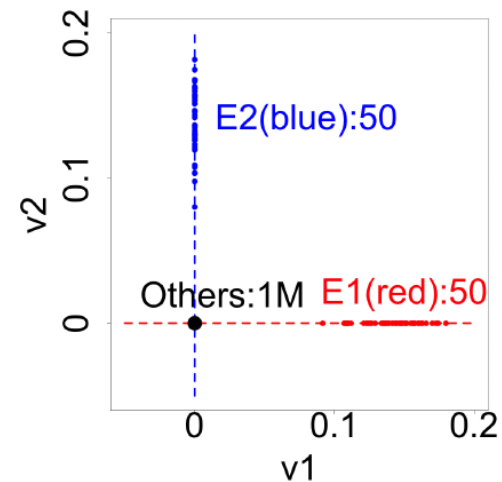
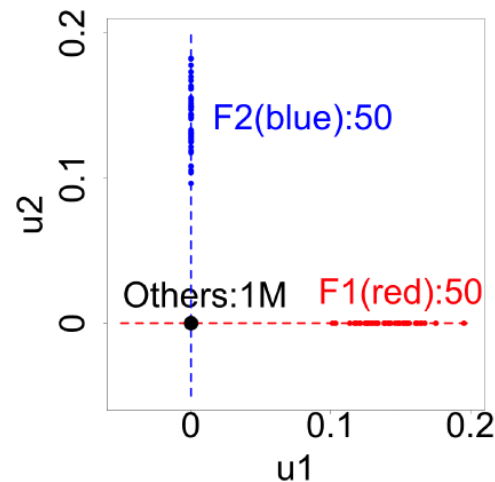
Spectral Subspace Plot: Case #2

❖ Non-overlapping: Low density, Elongation

Adjacency Matrix



Spectral Subspace Plot

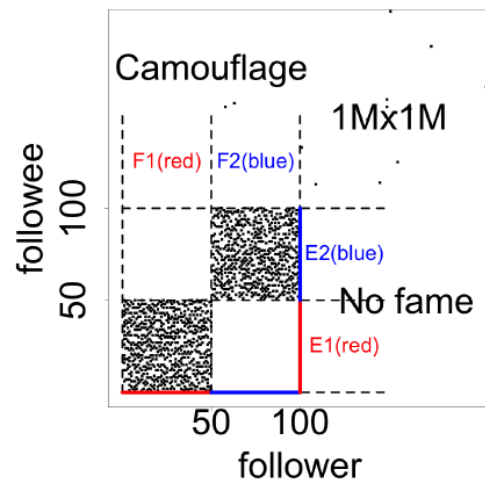


Rule 2 (long “rays”): two blocks, low density (50%), no “camouflage”, no “fame”

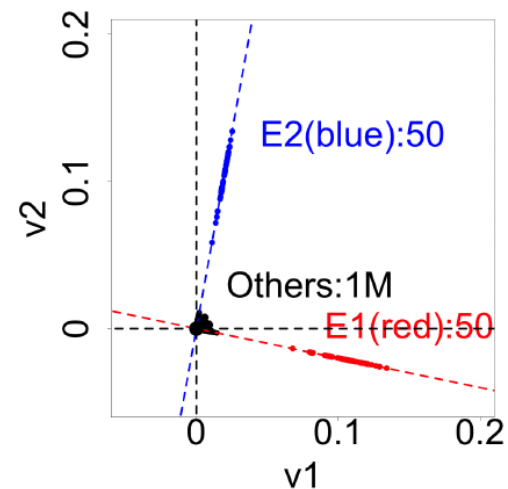
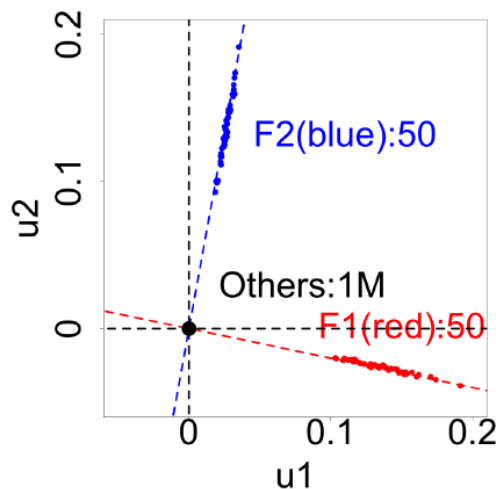
Spectral Subspace Plot: Case #3

❖ Non-overlapping: Camouflage/Fame, Tilting

Adjacency Matrix



Spectral Subspace Plot

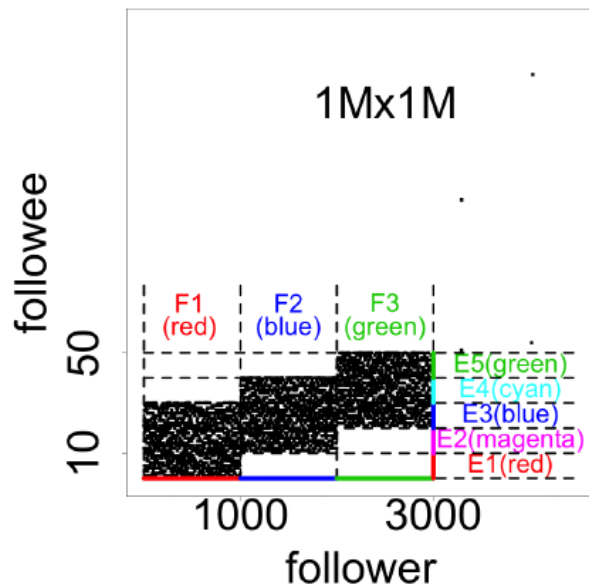


Rule 3 (tilting “rays”): two blocks, with “camouflage”, no “fame”

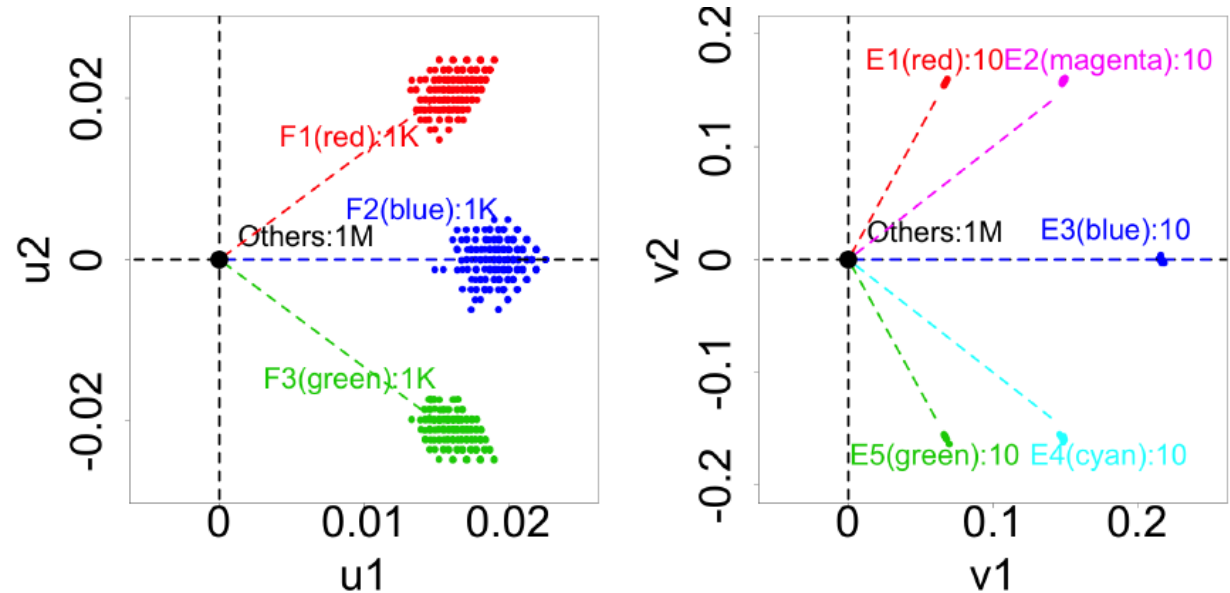
Spectral Subspace Plot: Case #4

❖ Overlapping: “Staircase”, “Pearls”

Adjacency Matrix



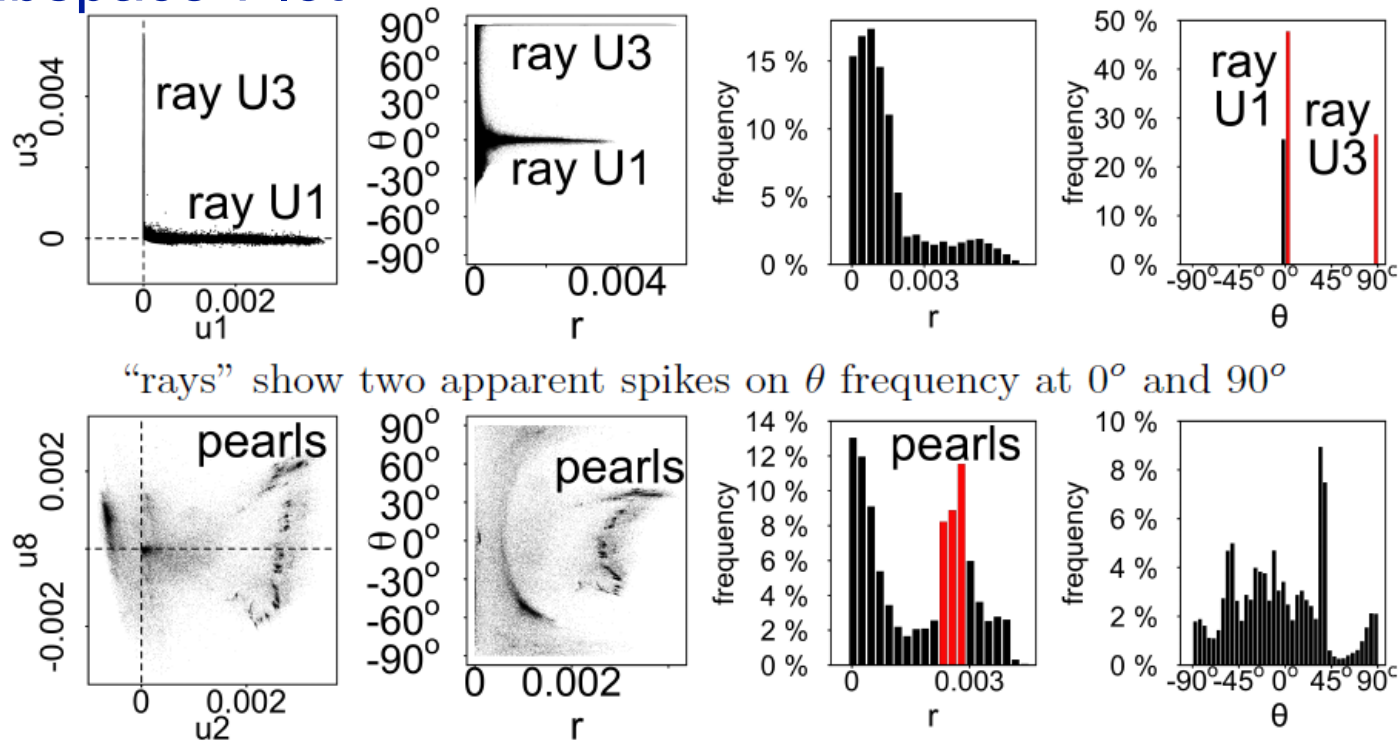
Spectral Subspace Plot



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

Spectral Subspace Plot: Reading & LockInfer

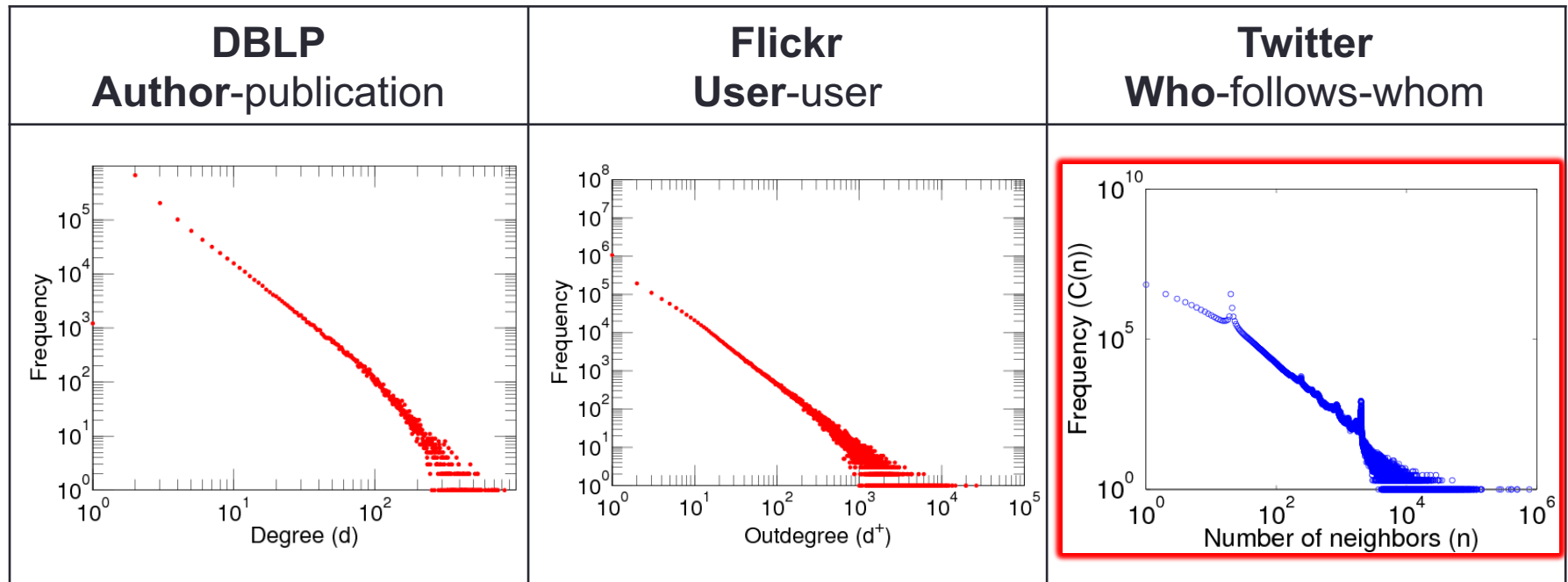
Spectral Subspace Plot Polar Coordinate Transform Histograms



“pearls” show a spike on r frequency at a much-greater-than-zero value

Graphical View: Power-Law Distribution

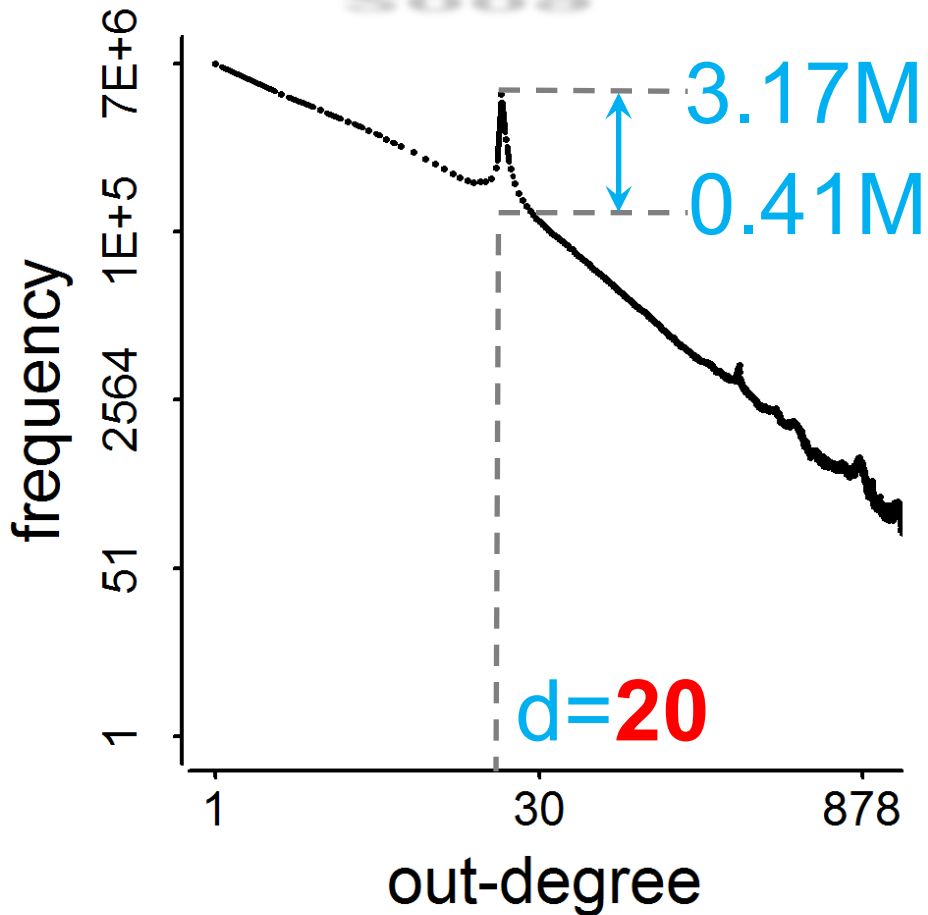
❖ Out-degree distribution



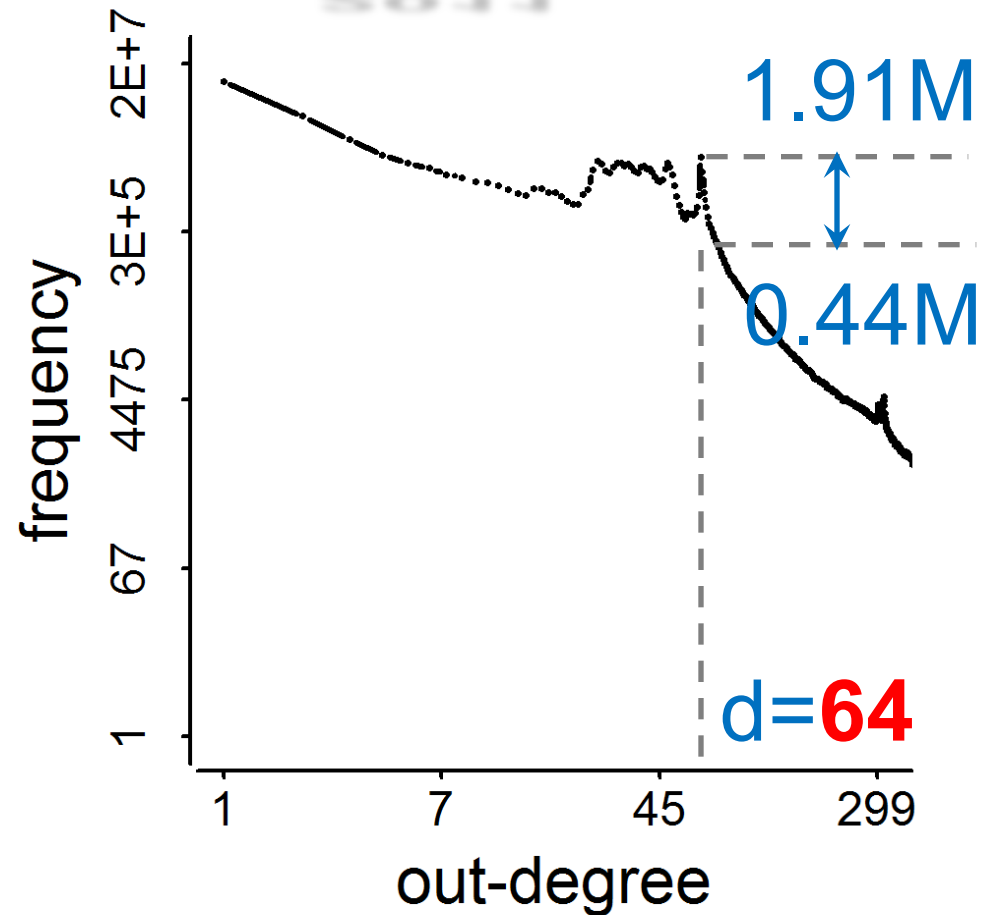
[konect.uni-koblenz.de/networks/]



2009 41M

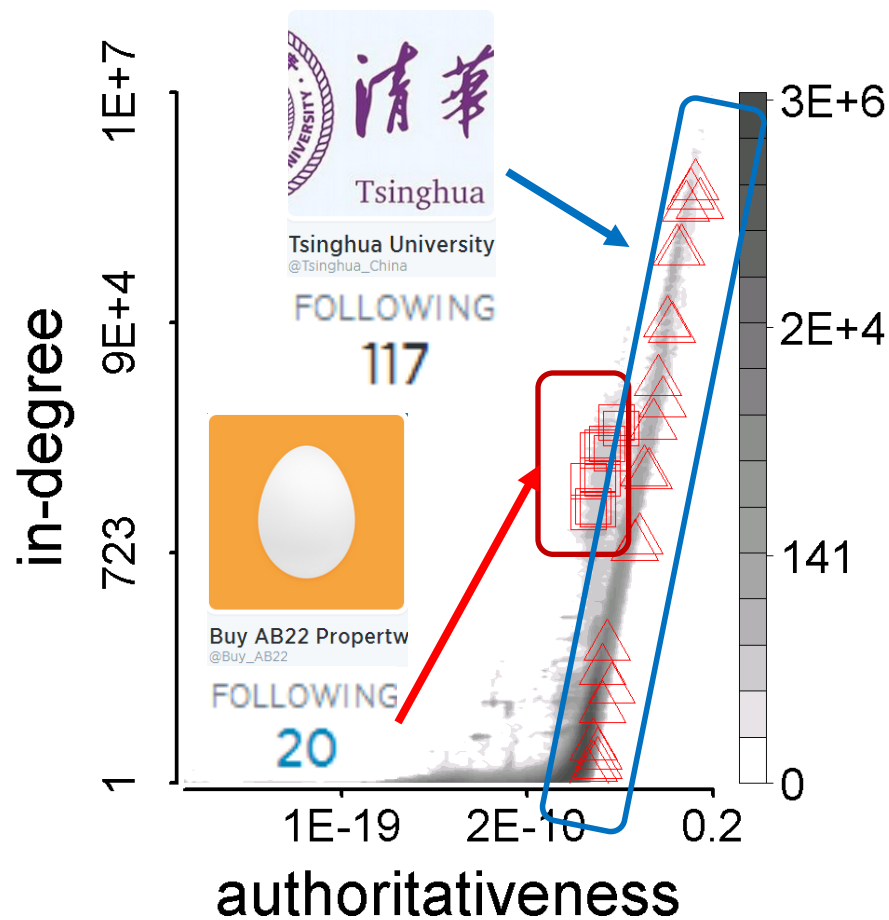


2011 117M



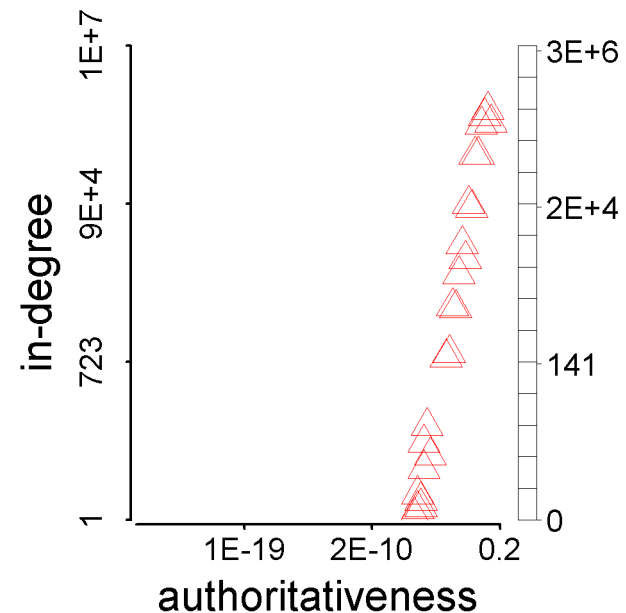
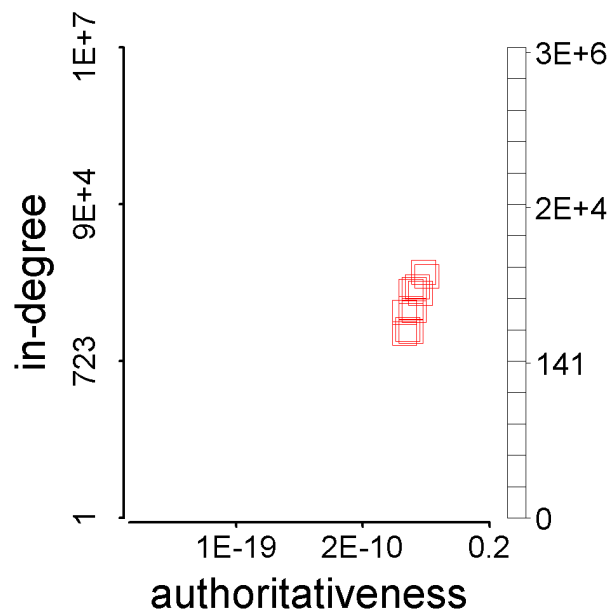
Synchronized Behavior: Features

- ❖ Synchronized
- ❖ Abnormal



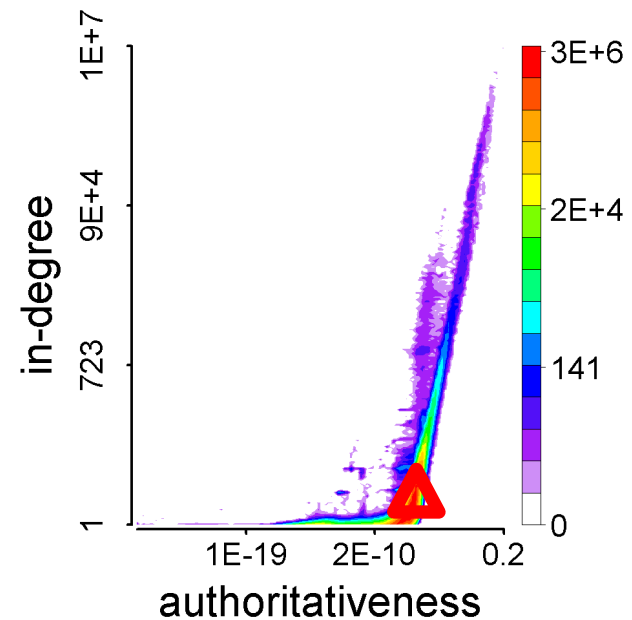
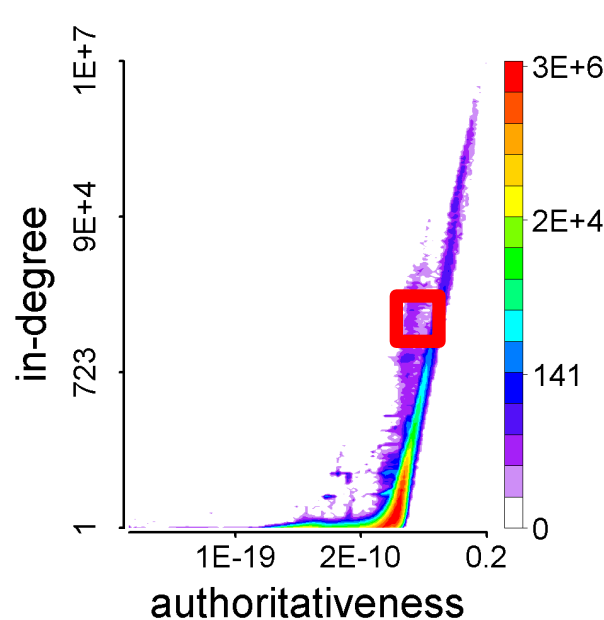
Synchronized Behavior: Synchronicity

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



Synchronized Behavior: Normality

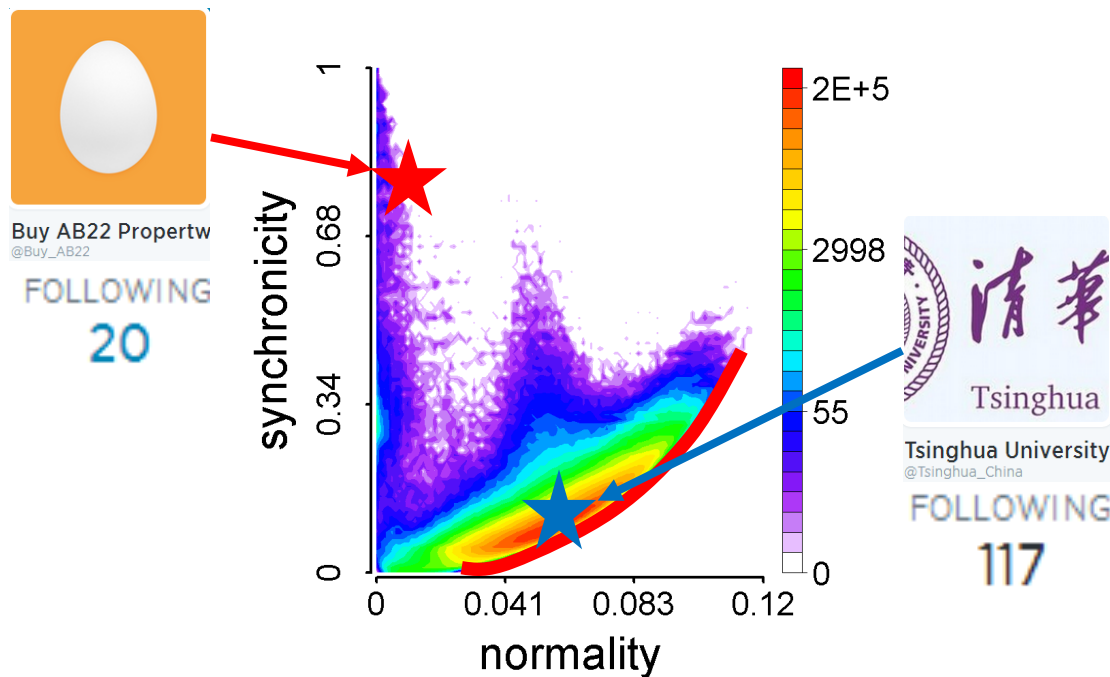
$$\text{norm}(u) = \frac{\sum_{(v,v') \in \mathcal{F}(u) \times \mathcal{U}} \mathbf{p}(v) \cdot \mathbf{p}(v')}{d(u) \times N}$$



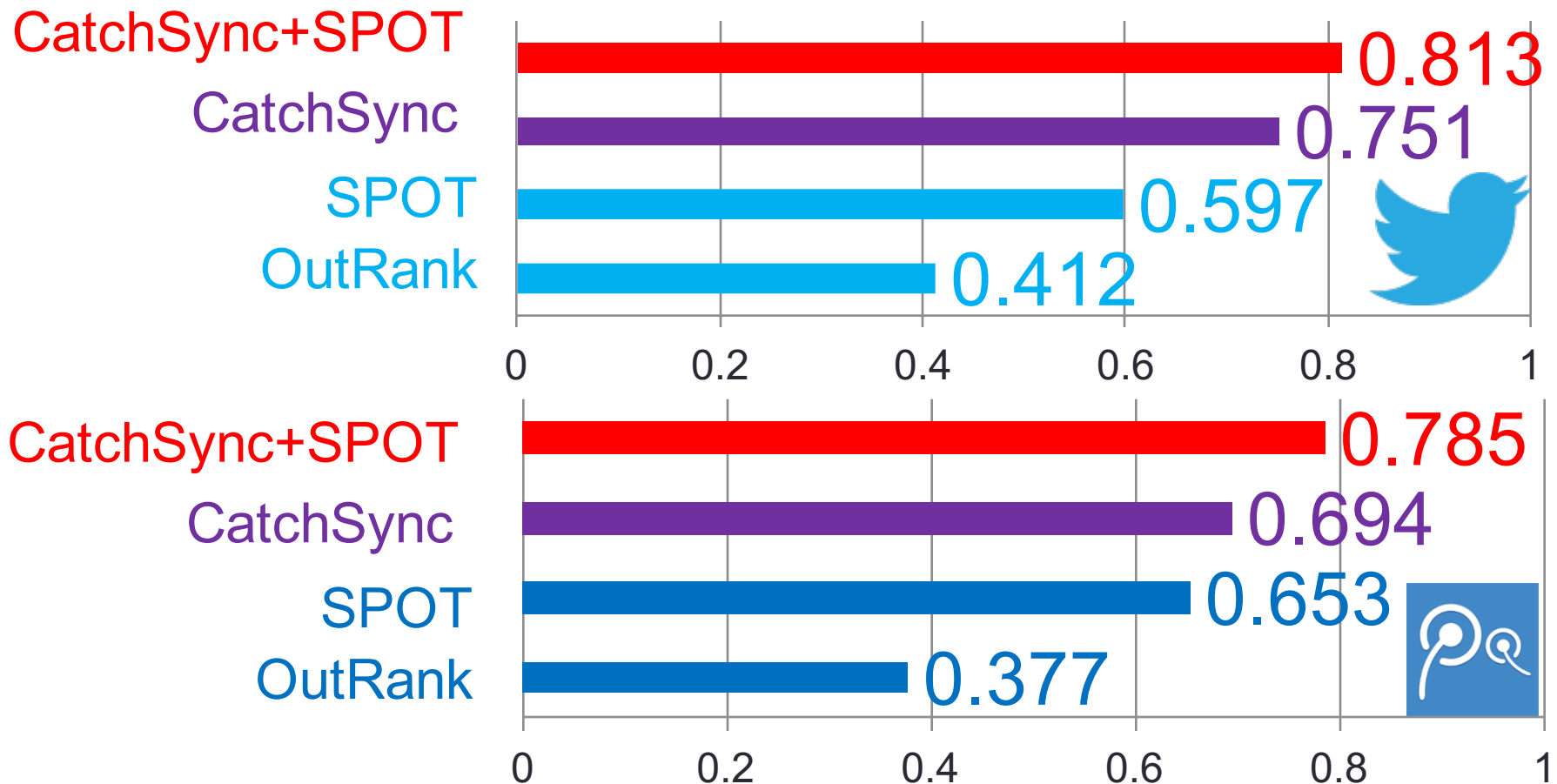
Synchronized Behavior: Theorem & CatchSync

❖ Synchronicity-Normality Plot

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



Synchronized Behavior: Performance

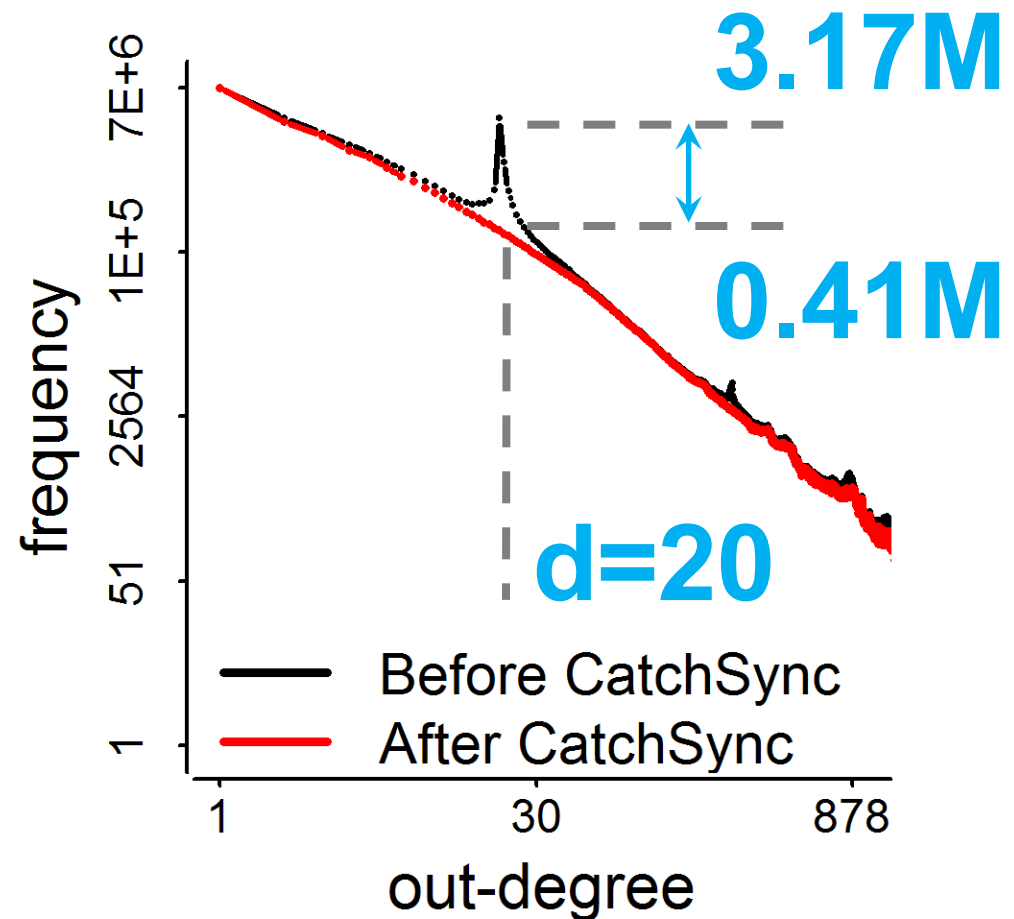
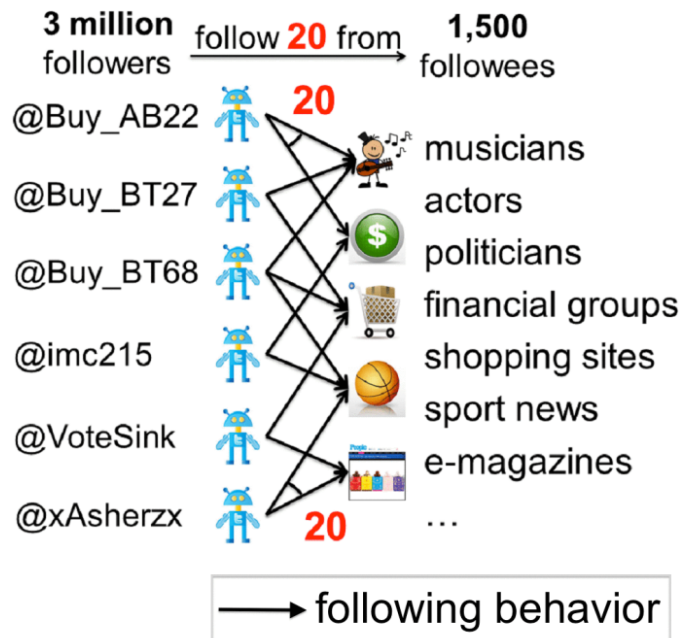


Synchronized Behavior: Performance

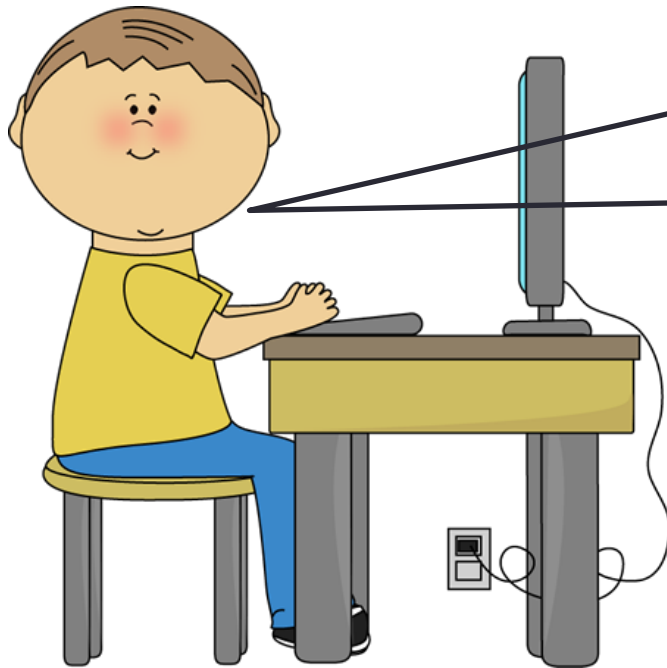


2009

41M



Beyond Graph: Multi-Dimensional Fraud

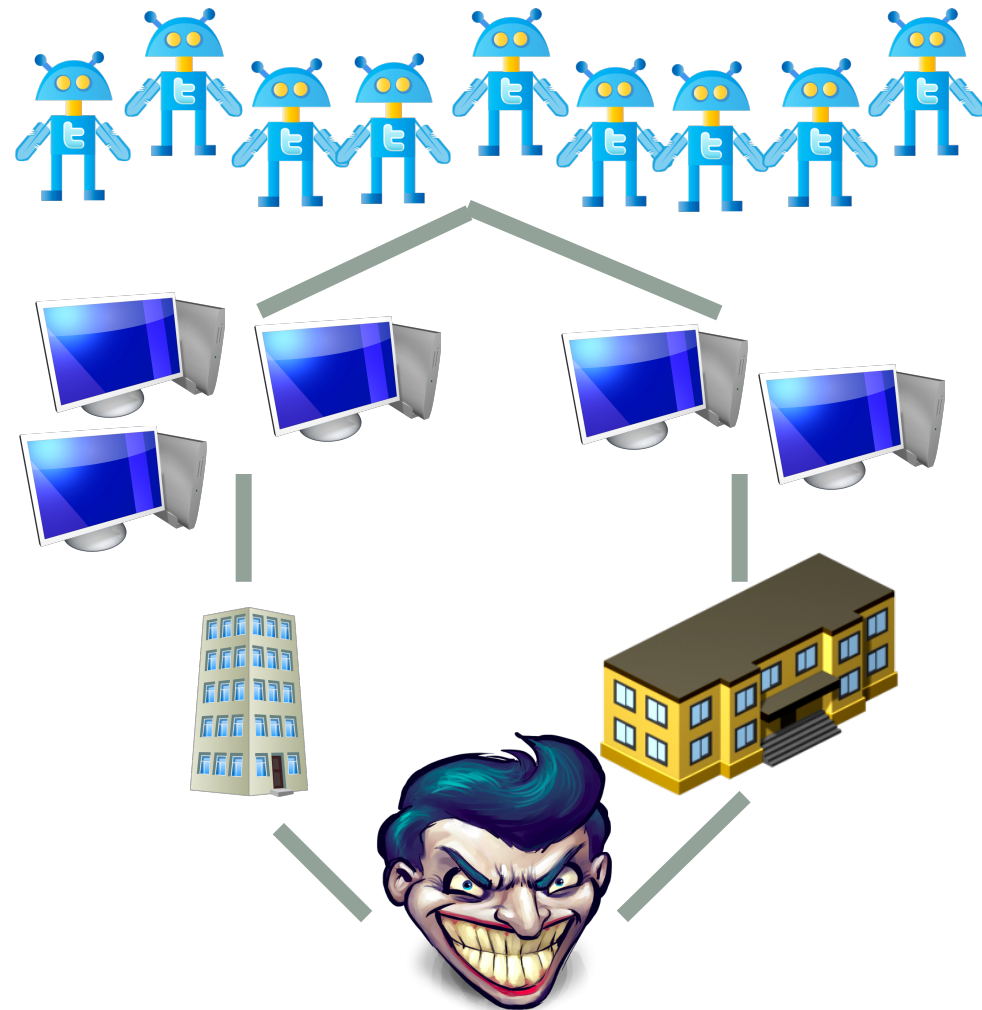
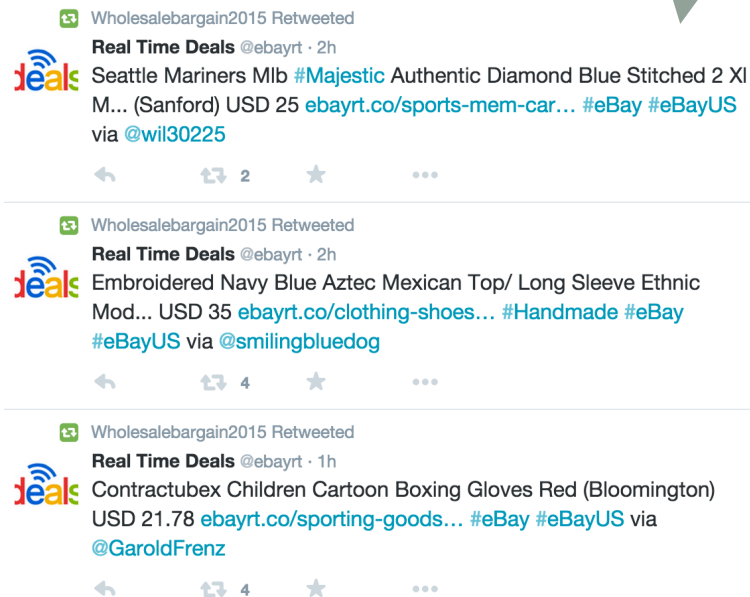


My boss wants me to **catch fraud** in such a big table – **billions of records, tens of columns!!!** How?!

	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. #dayyyym	dayyyym,
4	254	Nora Apnila J...	Dec 14, 2013	Alhamdulillahaaaaahhhh...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 evang 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

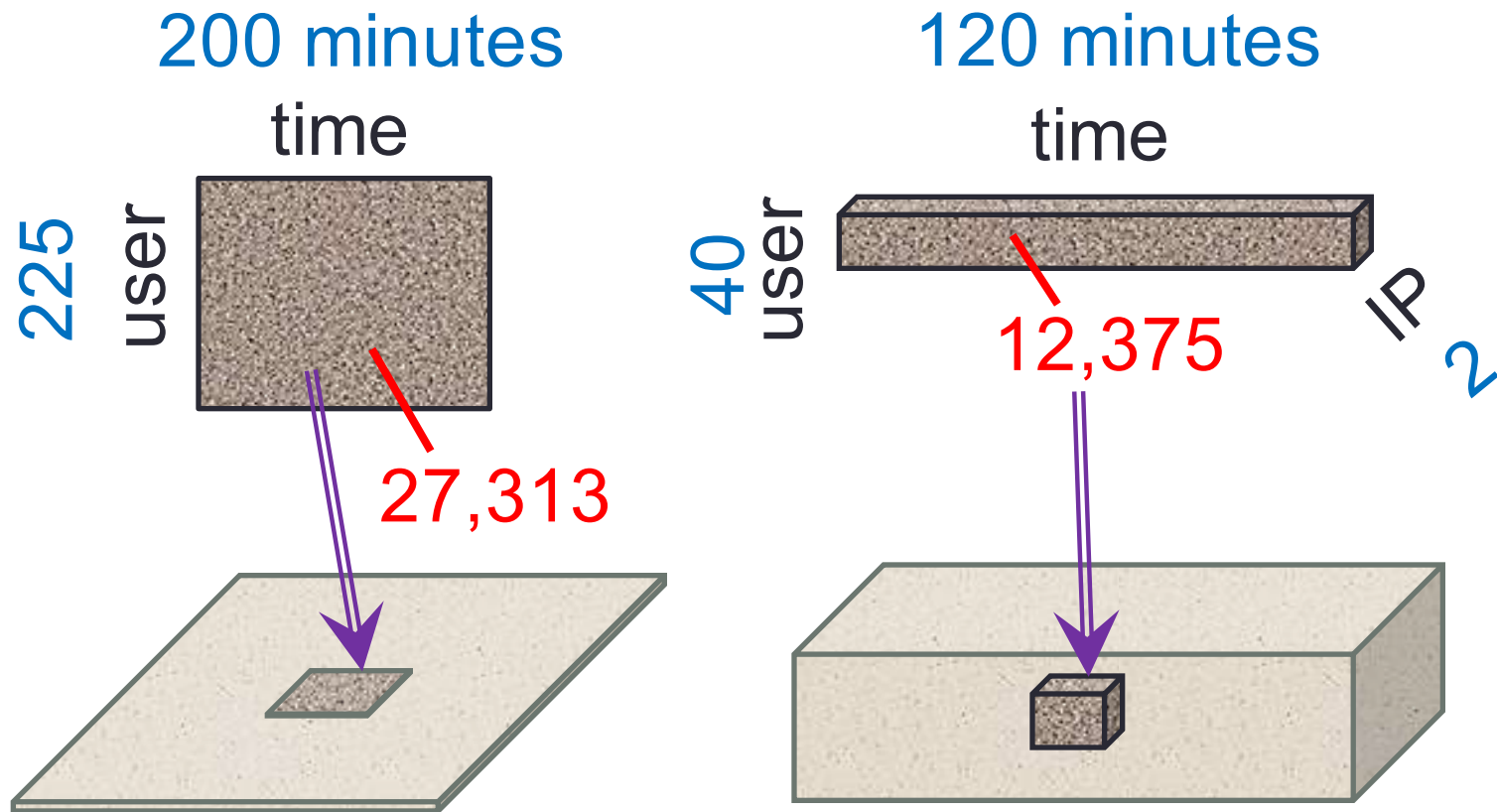
Multi-Dimensional Fraud



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

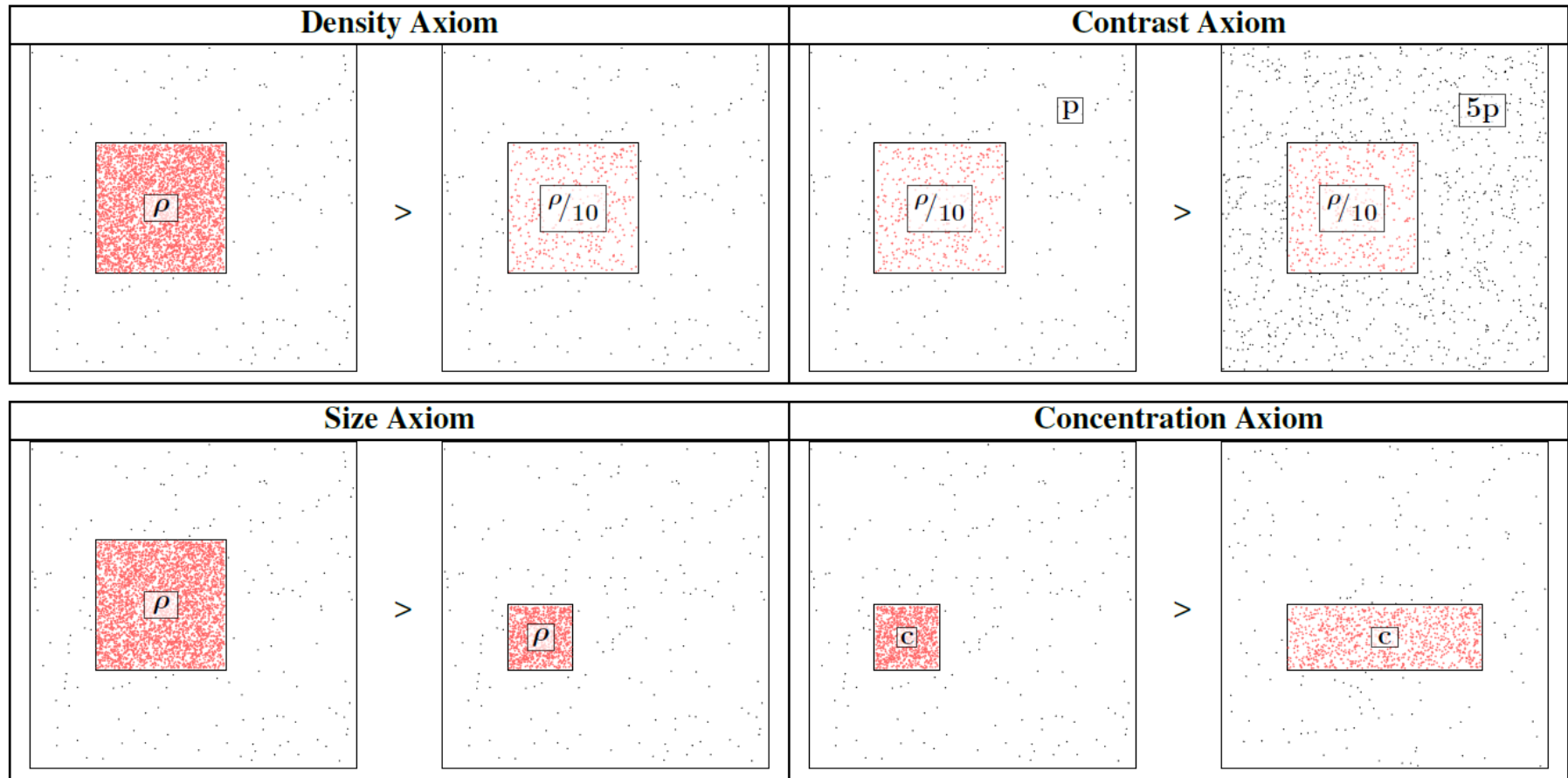
Suspiciousness: Density in Multi-Dimensions



Question: Which is more suspicious?

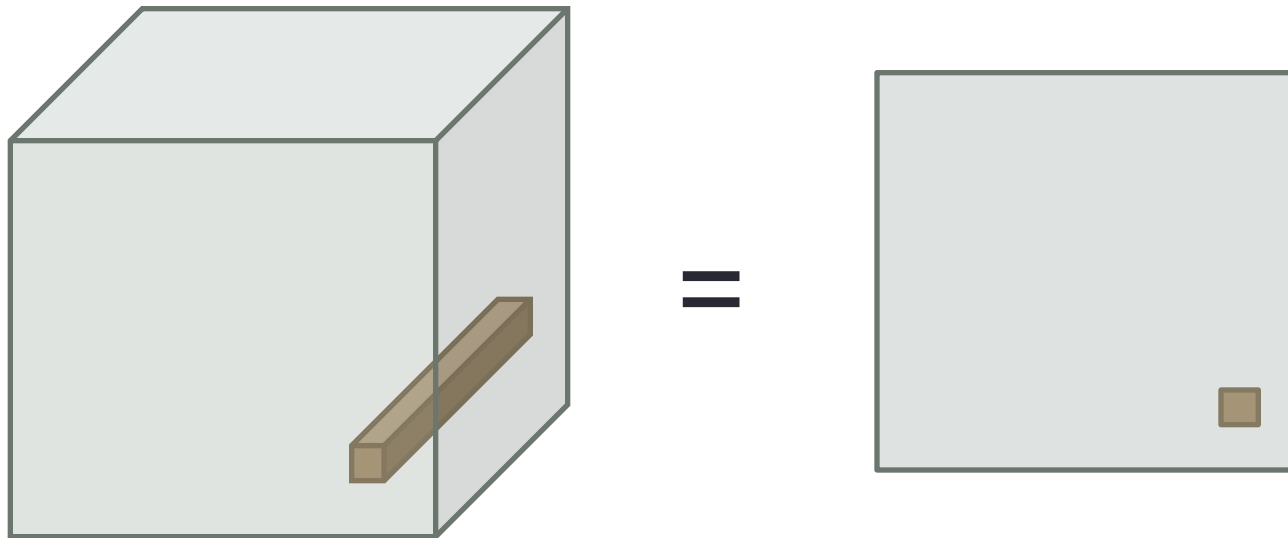
- Before we search, we should be able to rank.

Suspiciousness: Axiom 1 - 4



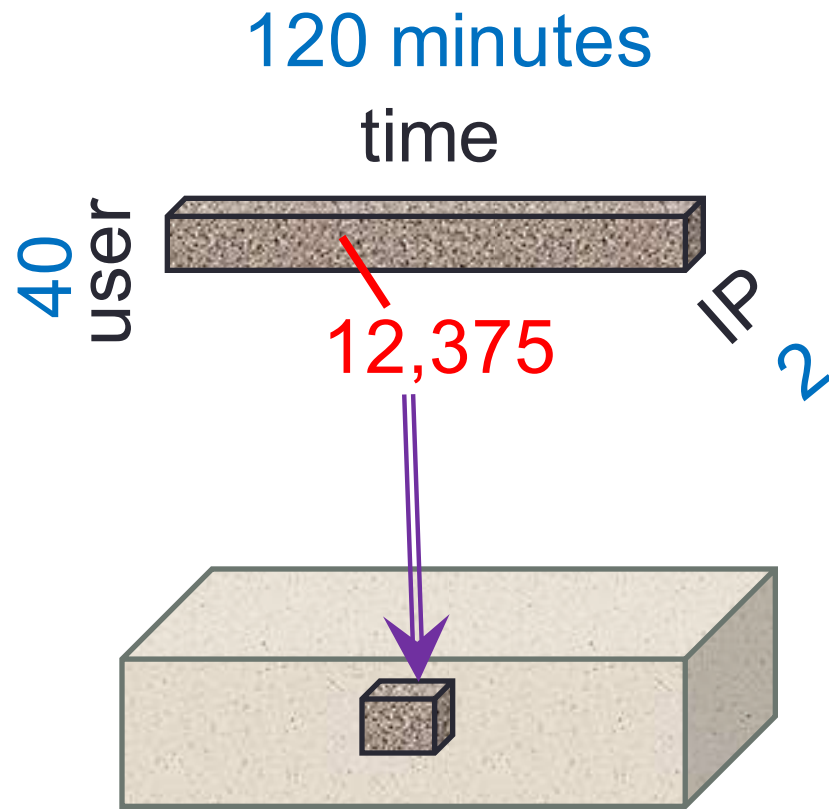
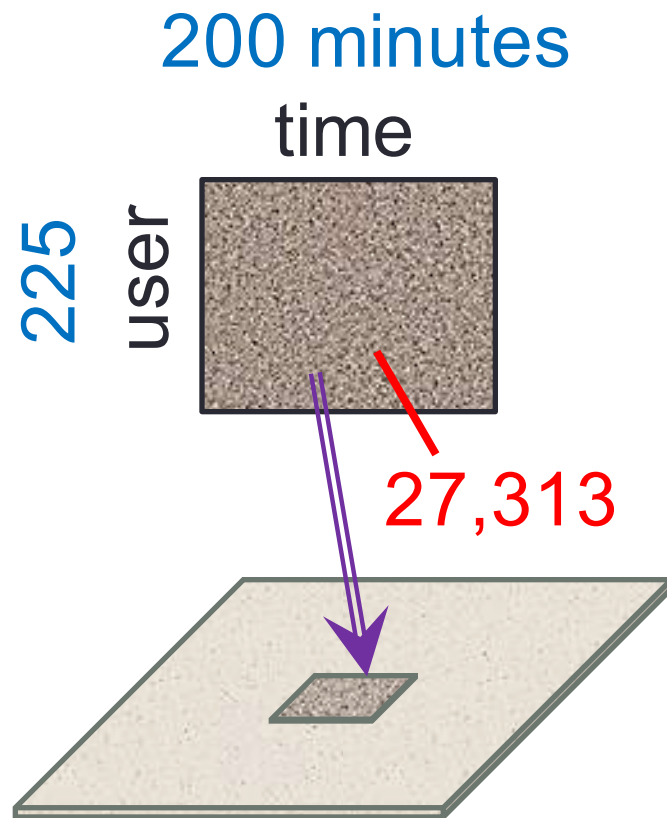
Suspiciousness: Axiom 5 Multimodal

Not including a mode is the same as including all values for that mode.

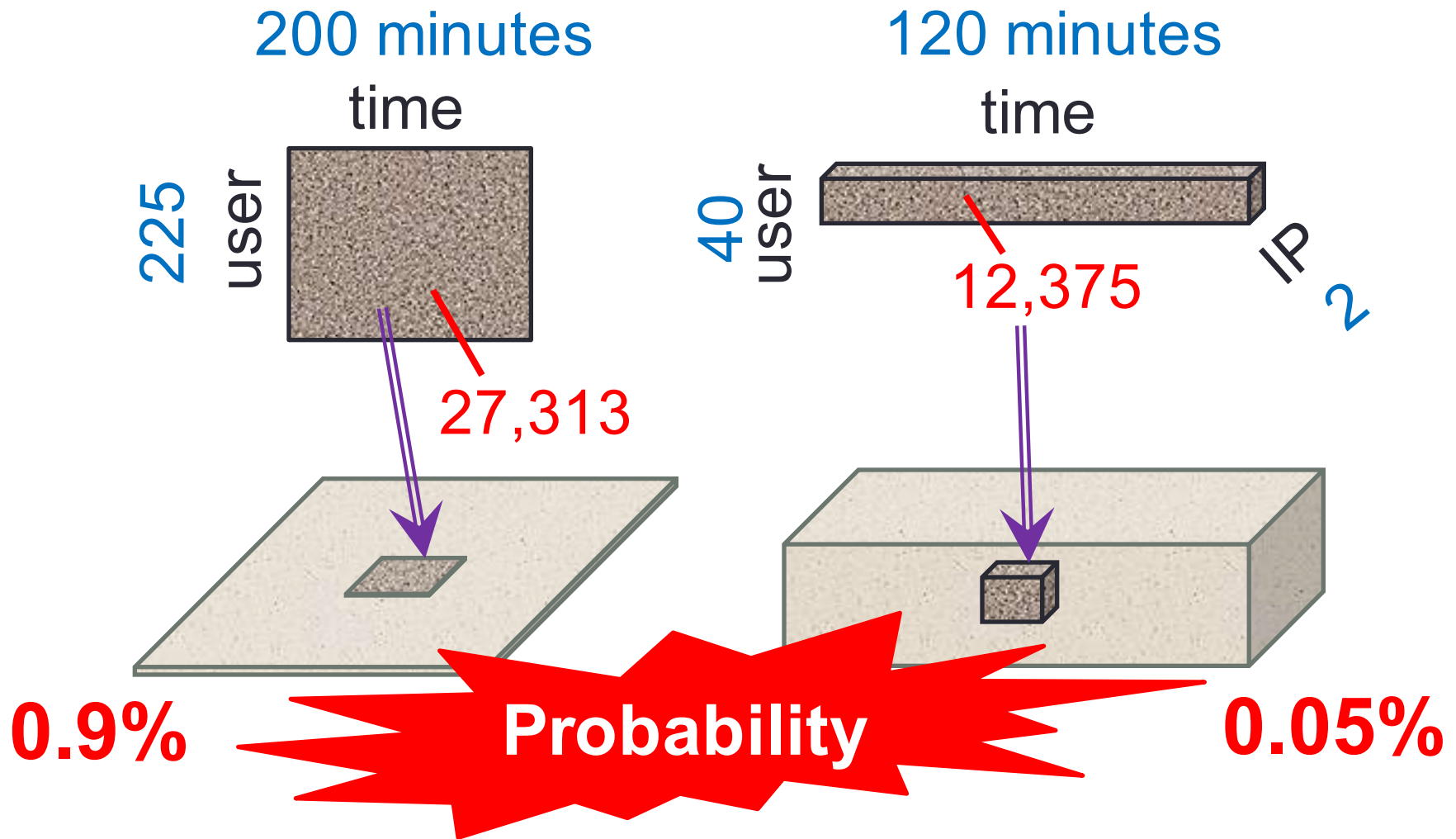


- ▶ New information (more modes) can only make our blocks more suspicious

Suspiciousness: General Metric & CrossSpot



Suspiciousness: General Metric & CrossSpot



Suspiciousness: General Metric & CrossSpot

❖ Negative log likelihood of block's probability

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

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

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

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

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

Suspiciousness: Trend Manipulating

User $\times$ hashtag $\times$ IP $\times$ minute	Mass c	Suspiciousness
$582 \times 3 \times 294 \times \mathbf{56,940}$	5,941,821	111,799,948
$188 \times 1 \times 313 \times \mathbf{56,943}$	2,344,614	47,013,868
$75 \times 1 \times 2 \times 2,061$	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!

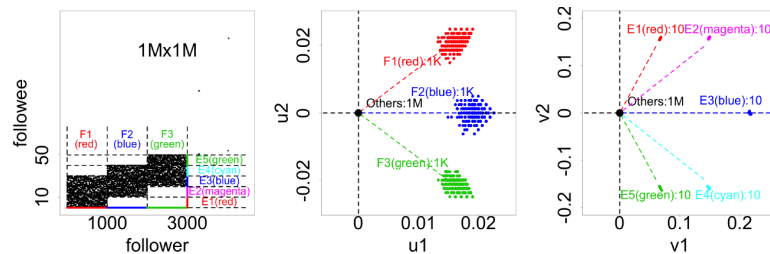
Summary for Detecting Unnatural Behavior

❖ Suspicious behavioral patterns

❖ **Lockstep pattern:** CopyCatch, LockInfer

❖ **Synchronized pattern:** CatchSync

❖ **Multi-dimensional suspiciousness:** CrossSpot



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

