KDD 2017
Halifax, Nova Scotia - Canada
August 13 - 17, 2017

Tutorial: Data-Driven Approaches towards Malicious Behavior Modeling

Meng Jiang
University of Notre Dame

Srijan Kumar
Stanford University

Christos Faloutsos
Carnegie Mellon University

V.S. Subrahmanian
University of Maryland, College Park

Tutorial link: http://bit.ly/kdd2017

Outline

Introduction

Feature-based algorithms

Sockpuppets

Vandals

Hoaxes

Spectral-based algorithms

Visualization: “spokes”, “blocks”, “staircases”

Camouflage

Theoretical guarantee

Density-based algorithms

Ill-gotten Likes

Synchronized Behaviors

Advertising campaigns

Social spam

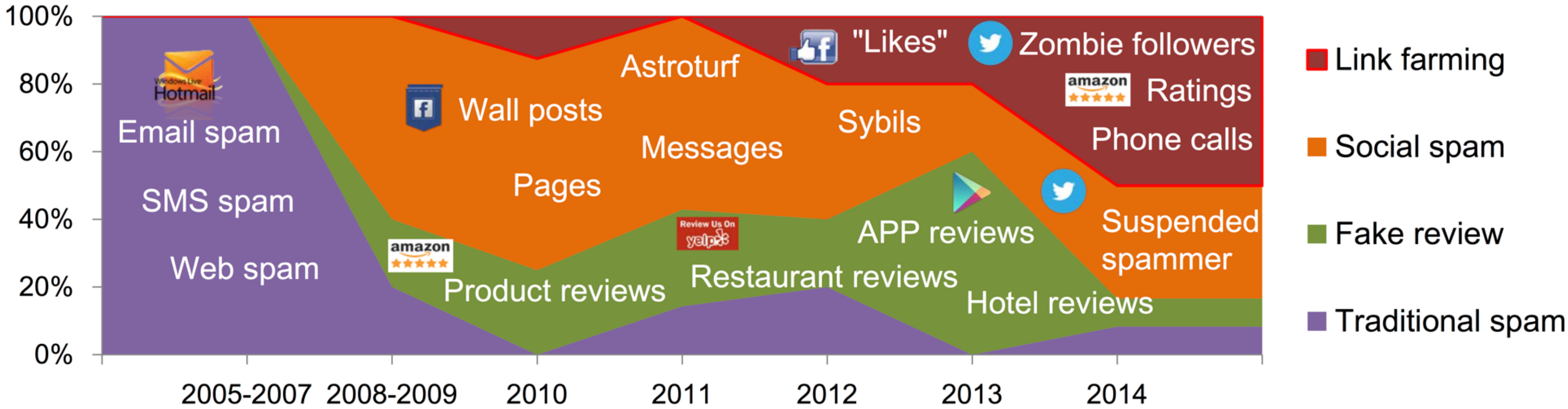
Conclusions and future directions

Tutorial link: http://bit.ly/kdd2017

Suspiciousness and Density

- What is worth of inspections in our data?
- Suspiciousness: tending to cause or excite suspicion
  - Unexpected high density
  - …
- Suspicious **density** of user behaviors in applications
  - Ill-gotten Likes: Facebook, etc.
  - Zombie followers: Twitter, Weibo, etc.
  - Social spam/fake reviews: Twitter, Weibo, Amazon, etc.
  - Advertising campaigns: Twitter, Weibo, etc.

… Social Spam, Social Link Farming

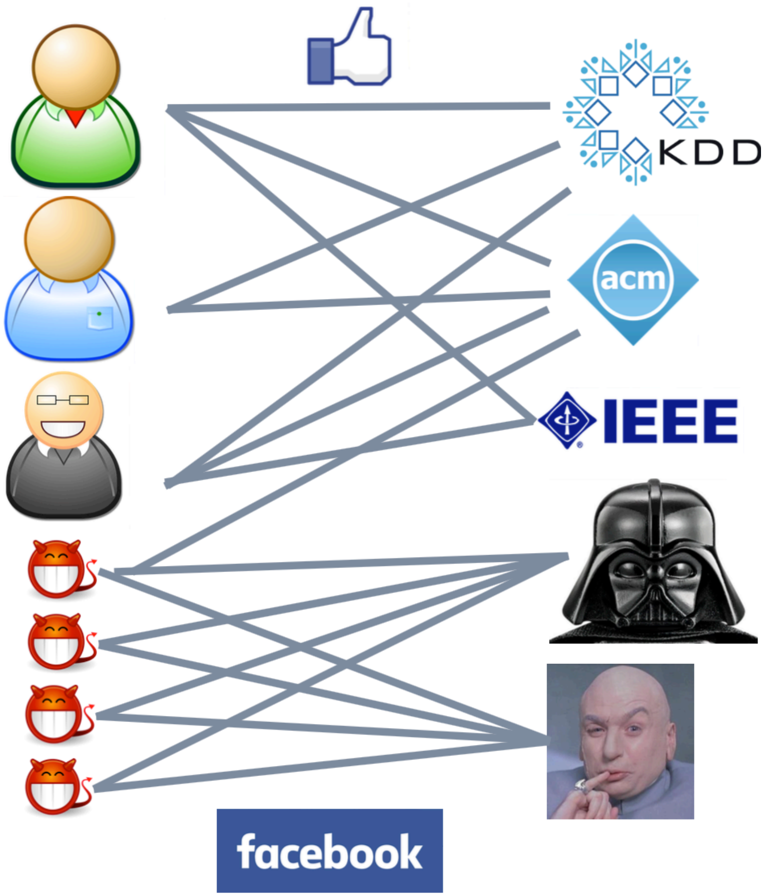


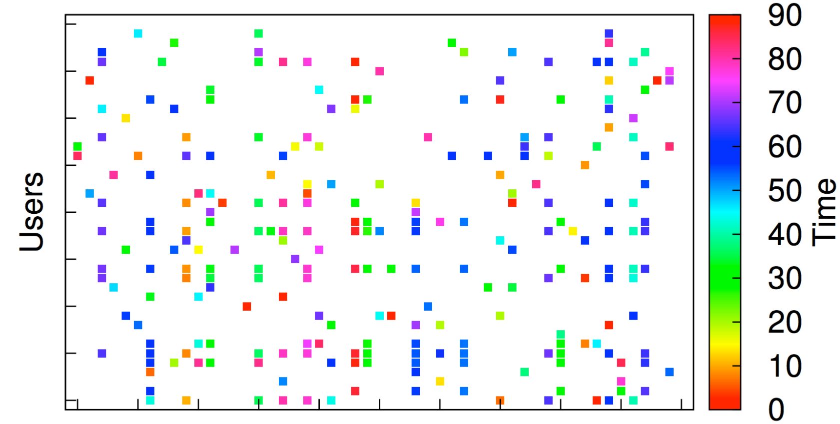
Meng Jiang, Peng Cui, and Christos Faloutsos. "Suspicious behavior detection: current trends and future directions." **IEEE Intelligent Systems**, 2016. (Survey paper)

1. Ill-gotten Facebook Likes

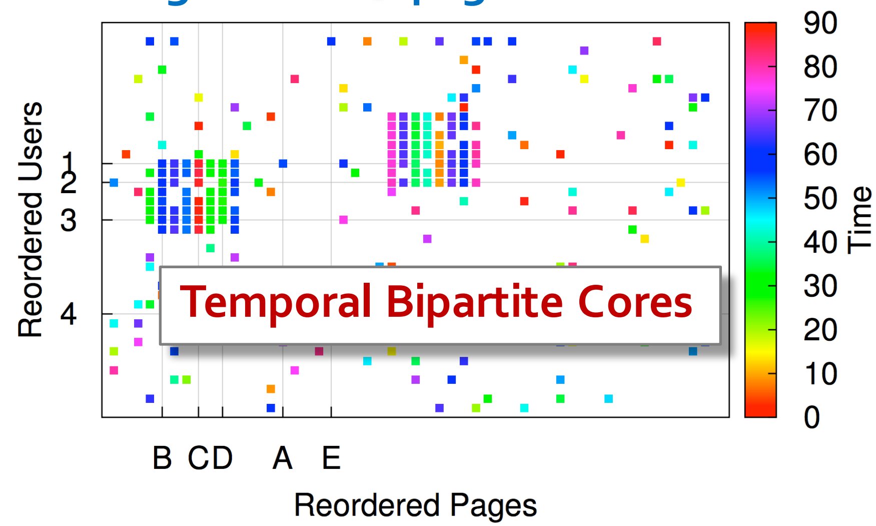
Alex Beutel, Wanhong Xu, Venkatesan Guruswami, Christopher Palow, Christos Faloutsos. “Copycatch: stopping group attacks by spotting lockstep behavior in social networks”, WWW 2013.

Density in Temporal Bipartite Graph





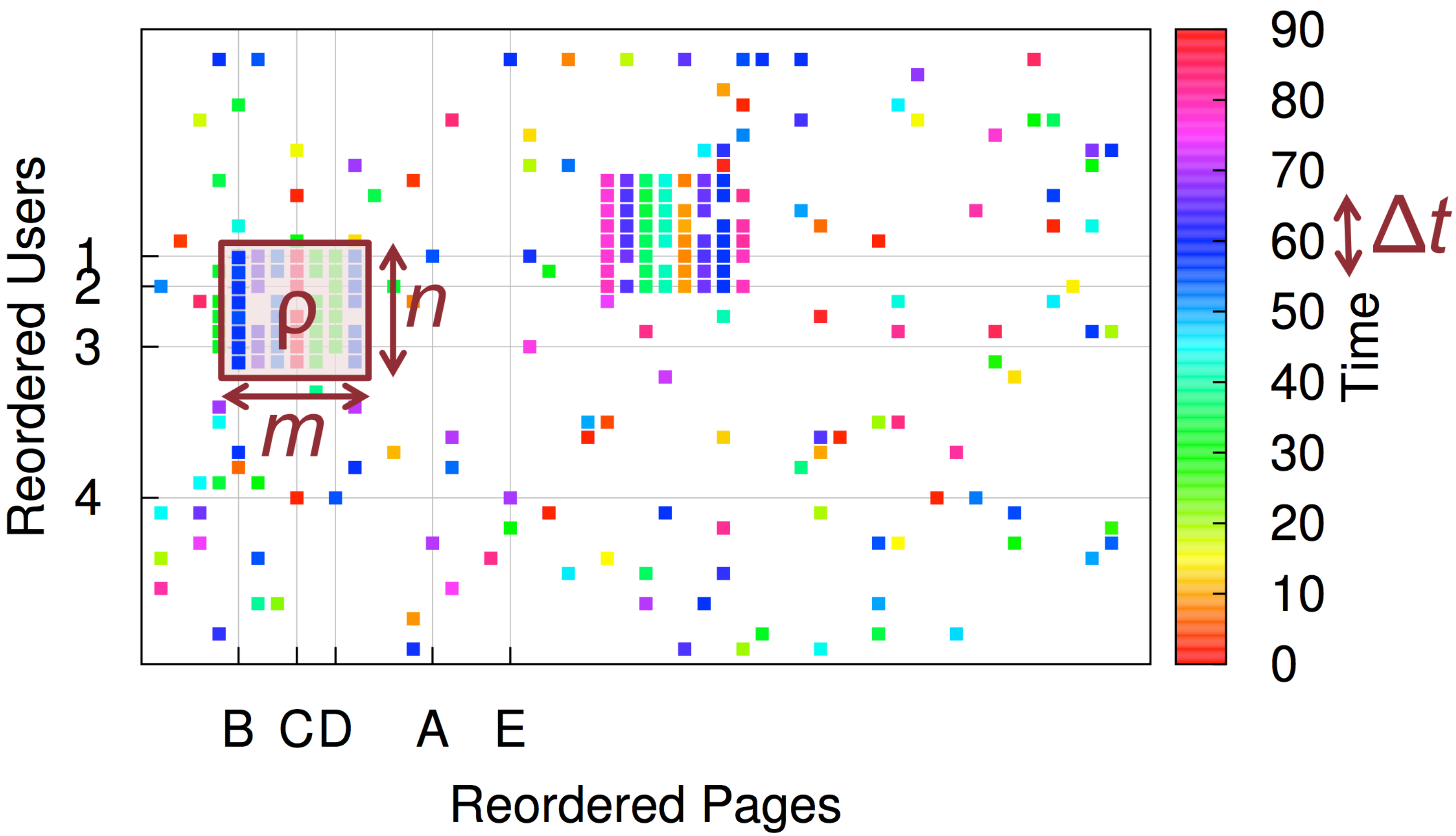
Reordering users and pages



Beutel, et al. (WWW 2013)

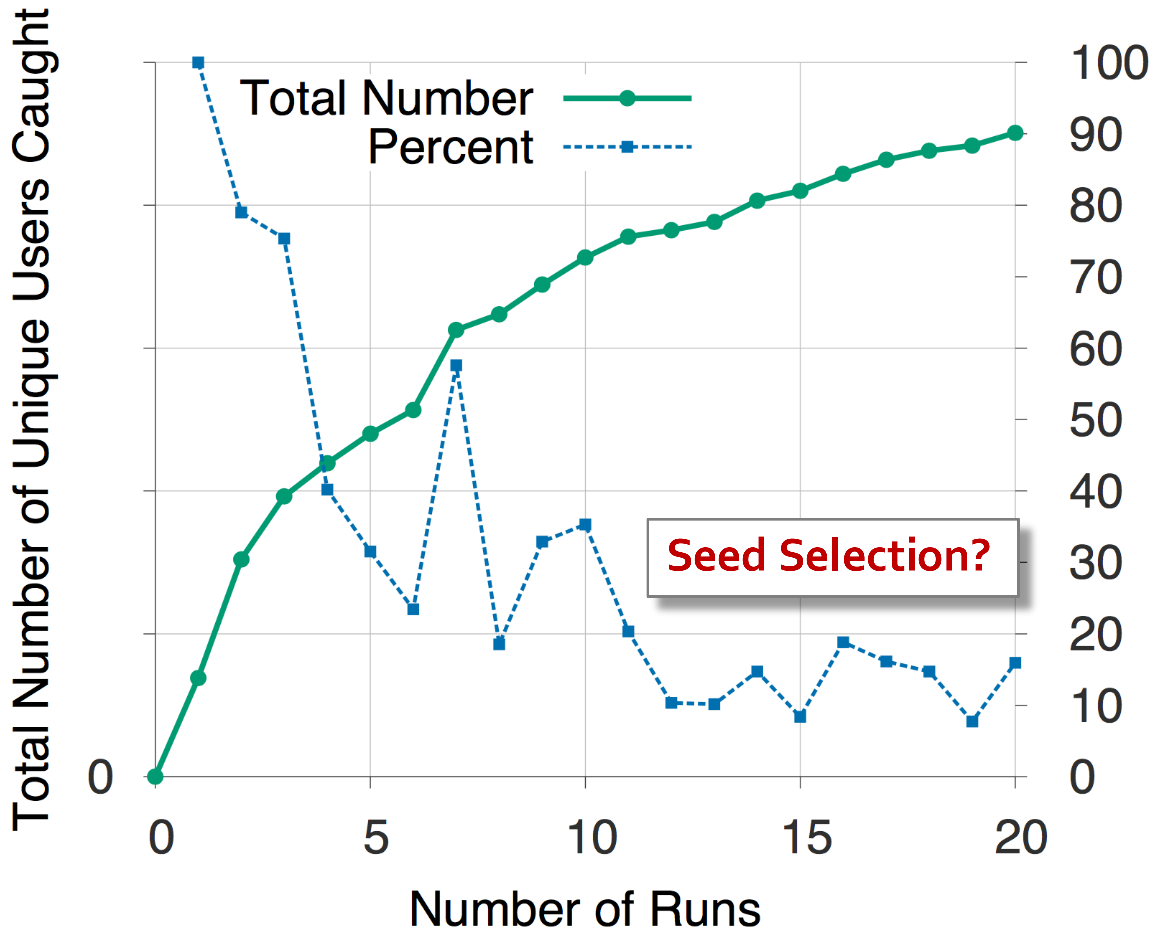
CopyCatch: Seed + Search

- "Temporal Bipartite Core": n users, m pages, ρ, Δt



Beutel, et al. (WWW 2013)

CopyCatch: Deployed in Facebook

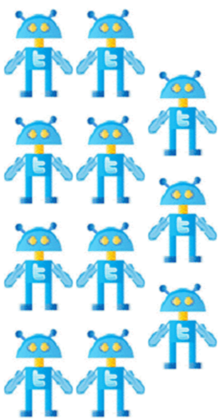


A: One option: Blocks/Staircases caught by LockInfer (spectral-based method).

2. Twitter's Zombie Followers

Meng Jiang, Peng Cui, Alex Beutel, Christos Faloutsos, Shiqiang Yang. "CatchSync: Catching Synchronized Behavior in Large Directed Graphs", KDD 2014 **Best Paper Finalist**.

Catching Zombie Followers



"Your best followers!"
"Delivery within 1-2 days"

engineers

product managers

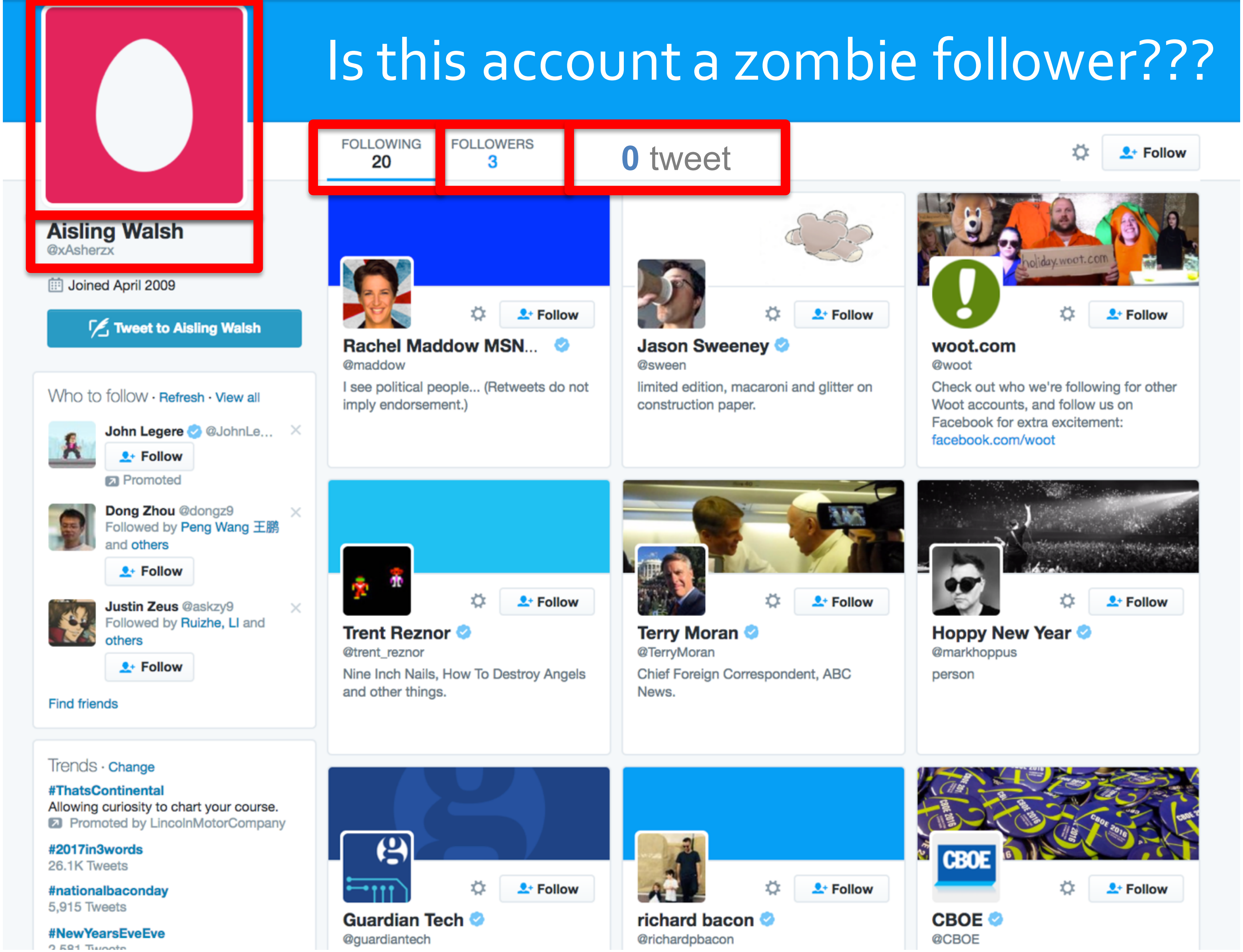
Knowledge from manual inspection: #followees, #followers, #tweets, #hashtags, #urls…

Learning models (classifiers)

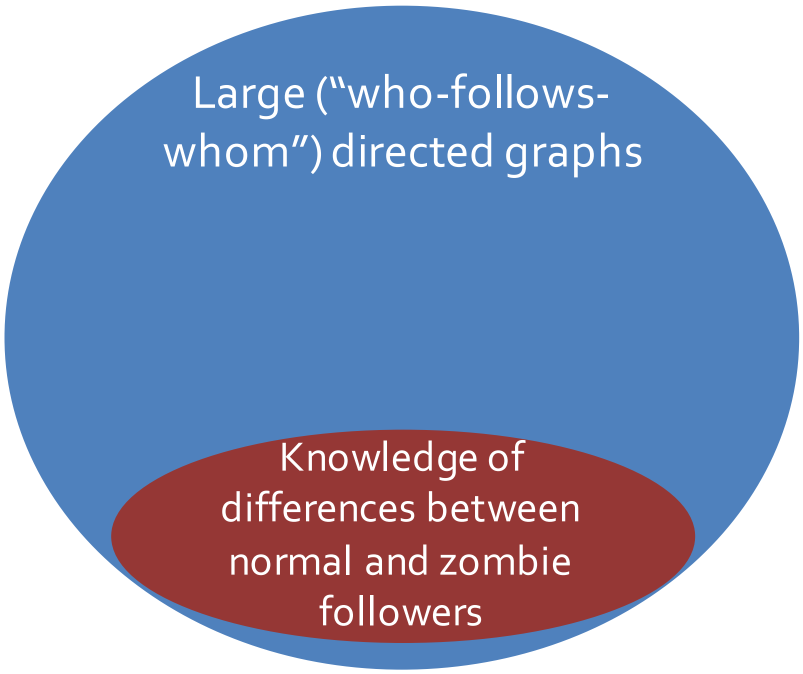
Poor accuracy
(**serious complaints** from users)

Fake account detection [Egele and Stringhini et al. NDSS'13; Yang and Wilson et al. TKDD'14; Viswanath and Bashir et al. USENIX Security Symposium'14]

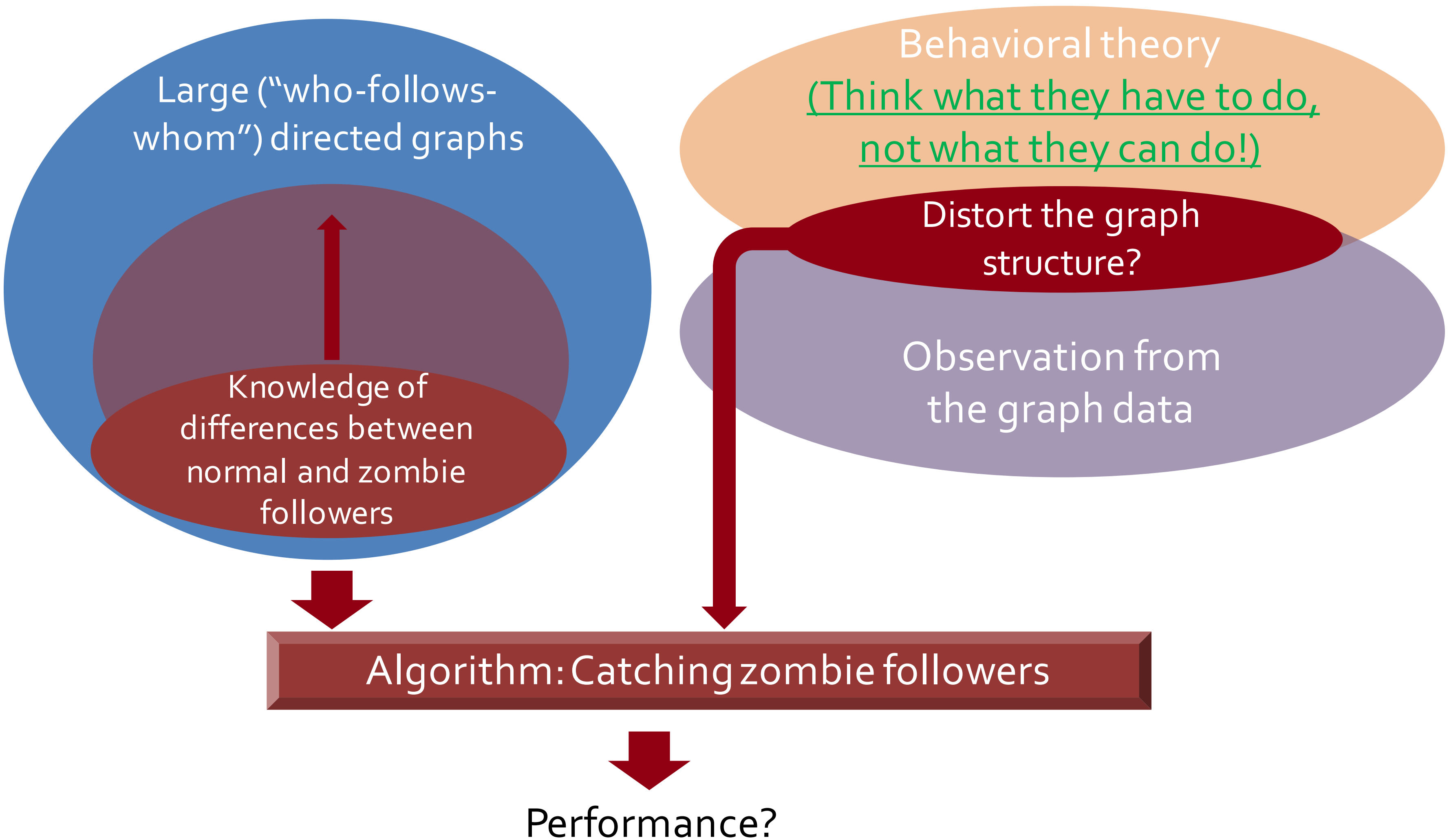
Is this account a zombie follower???



"Density" in Directed Graph

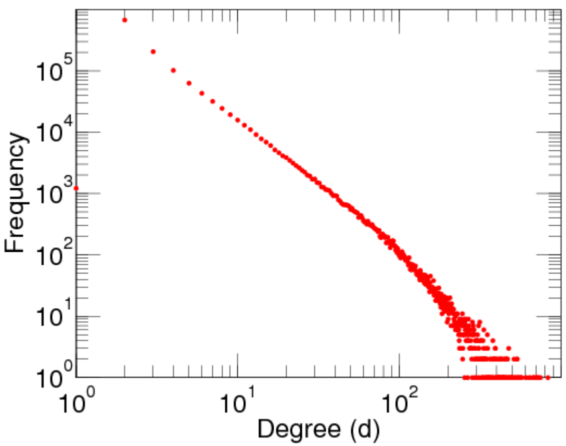


“Density” in Directed Graph

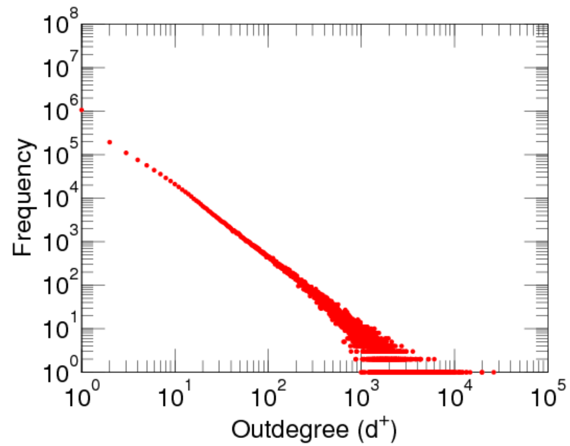


Out-Degree Distributions: Power Law Expected

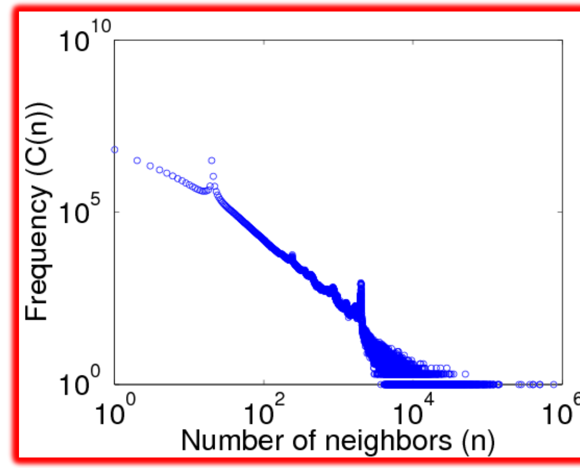
DBLP
Author-publication



Flickr
User-user



Twitter
Who-follows-whom



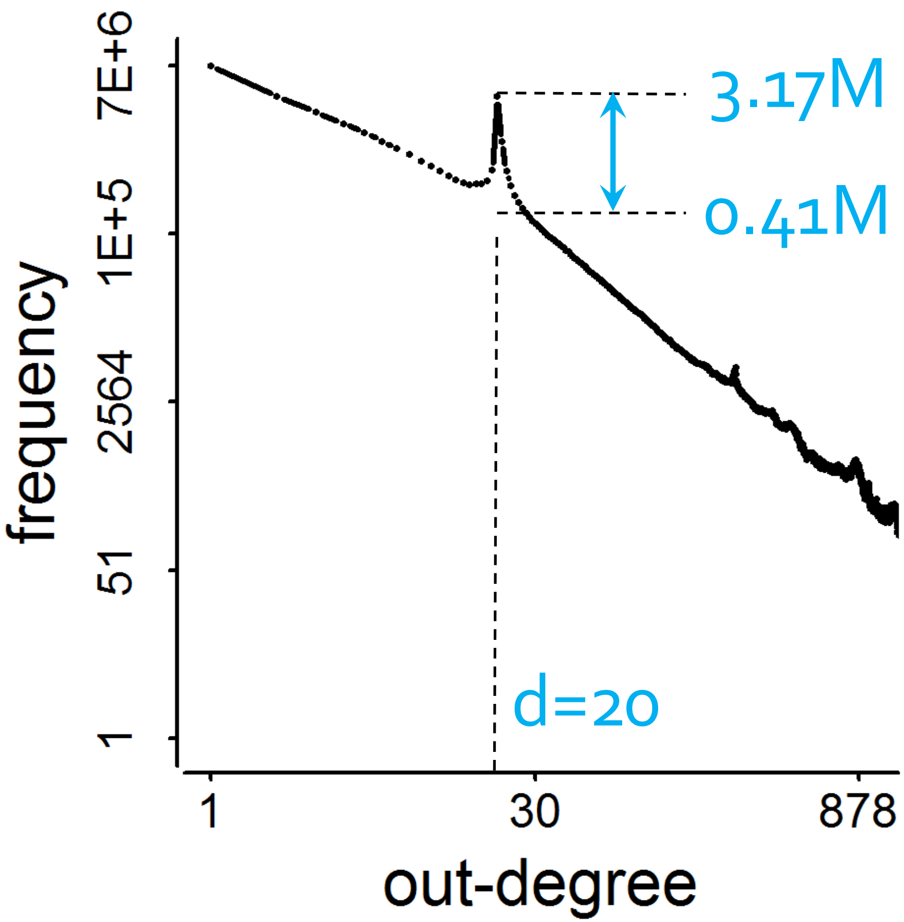
[konect.uni-koblenz.de/networks/]

Power-law distributions in networks

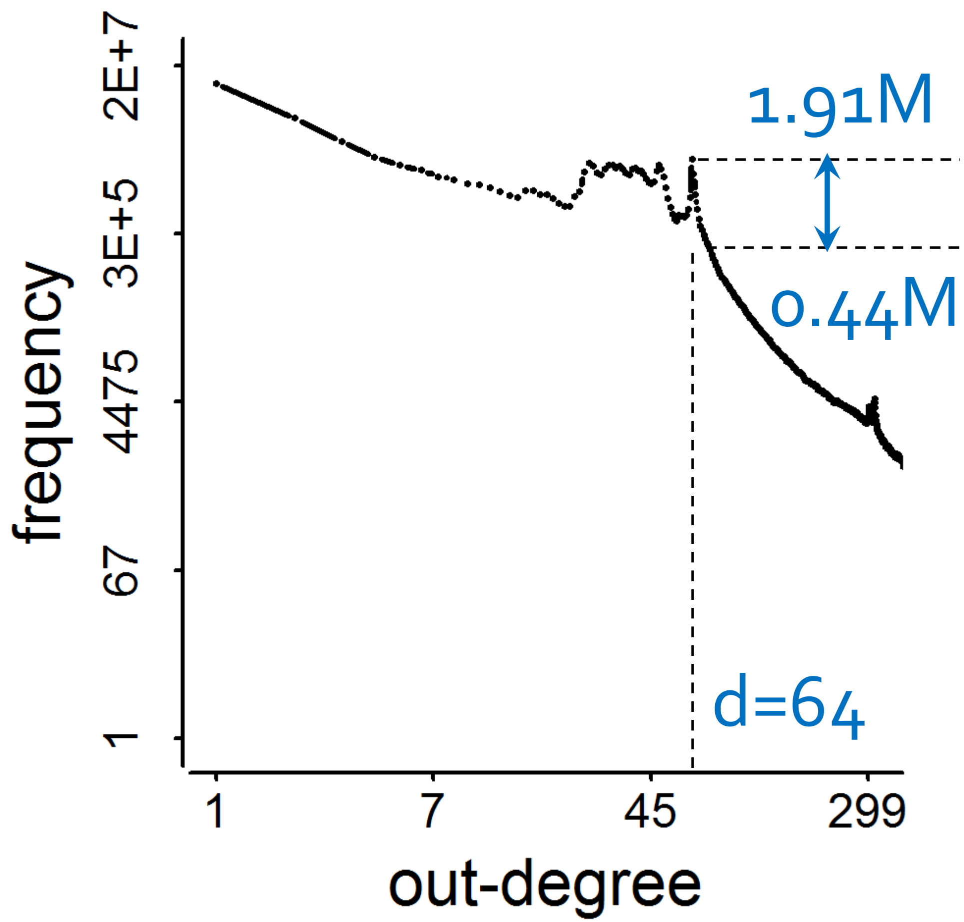
[Faloutsos et al. SIGCOMM'99; Chung et al. PNAS'02]

Spikes!

2009 41M

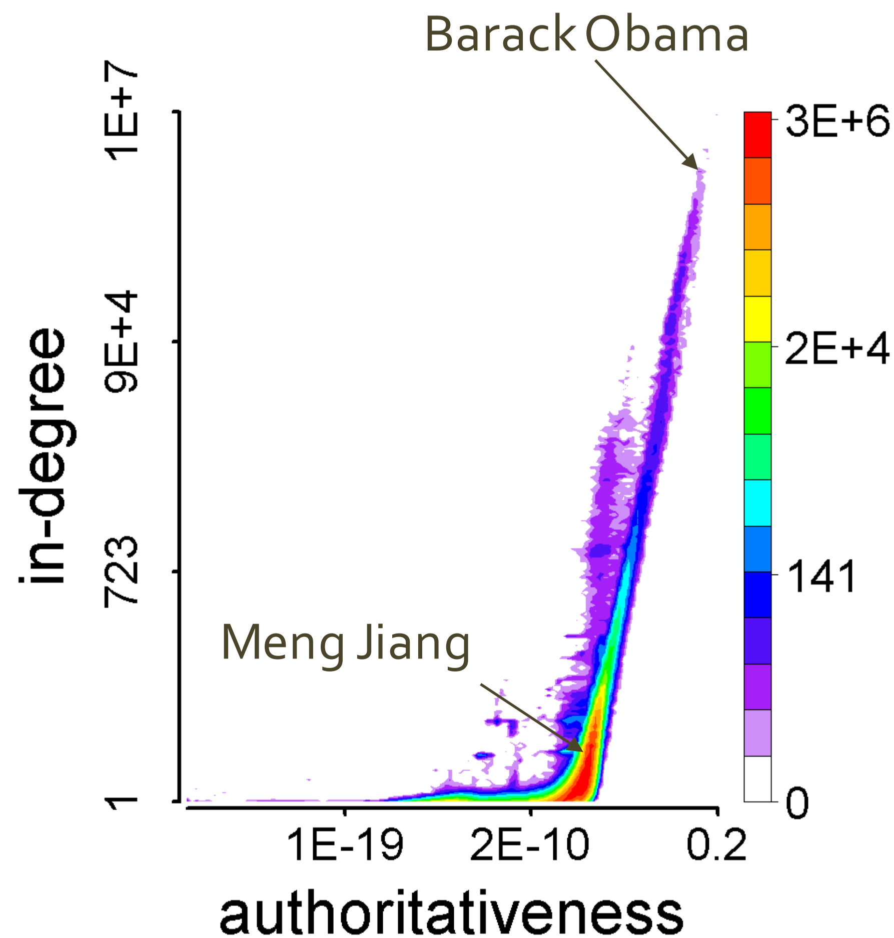


2011 117M



Jiang, et al. (KDD 2014)

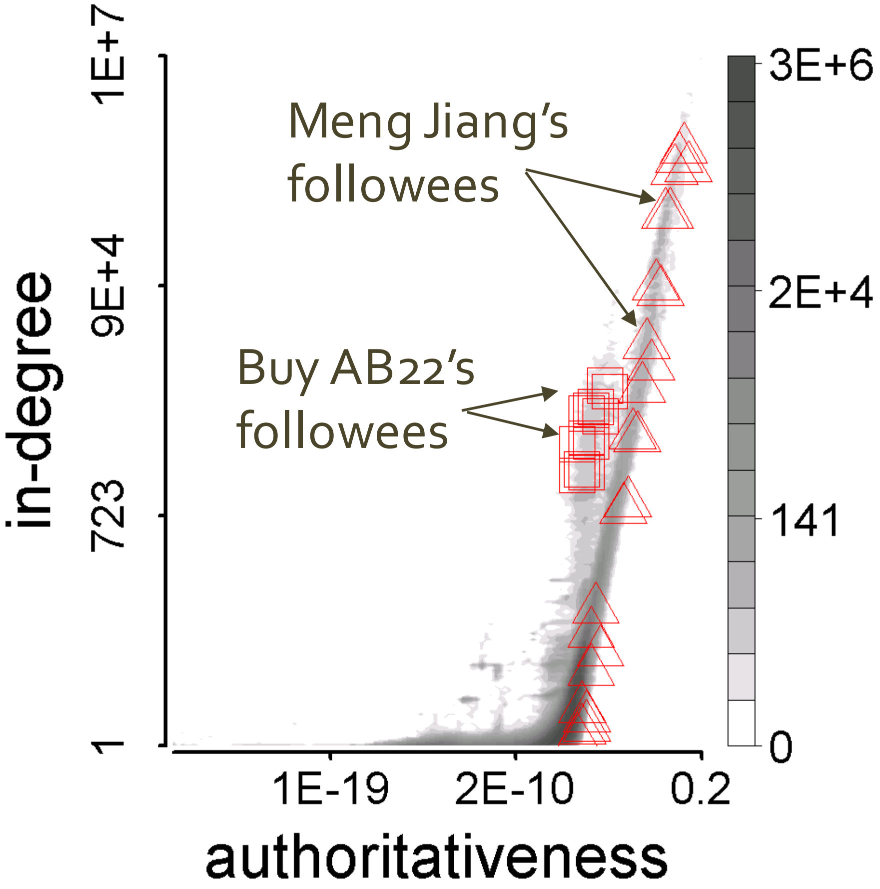
How We/They Connect to Our/Their Followees



The HITS algorithm. Kleinberg. “Authoritative sources in a hyperlinked environment.” JACM’99.

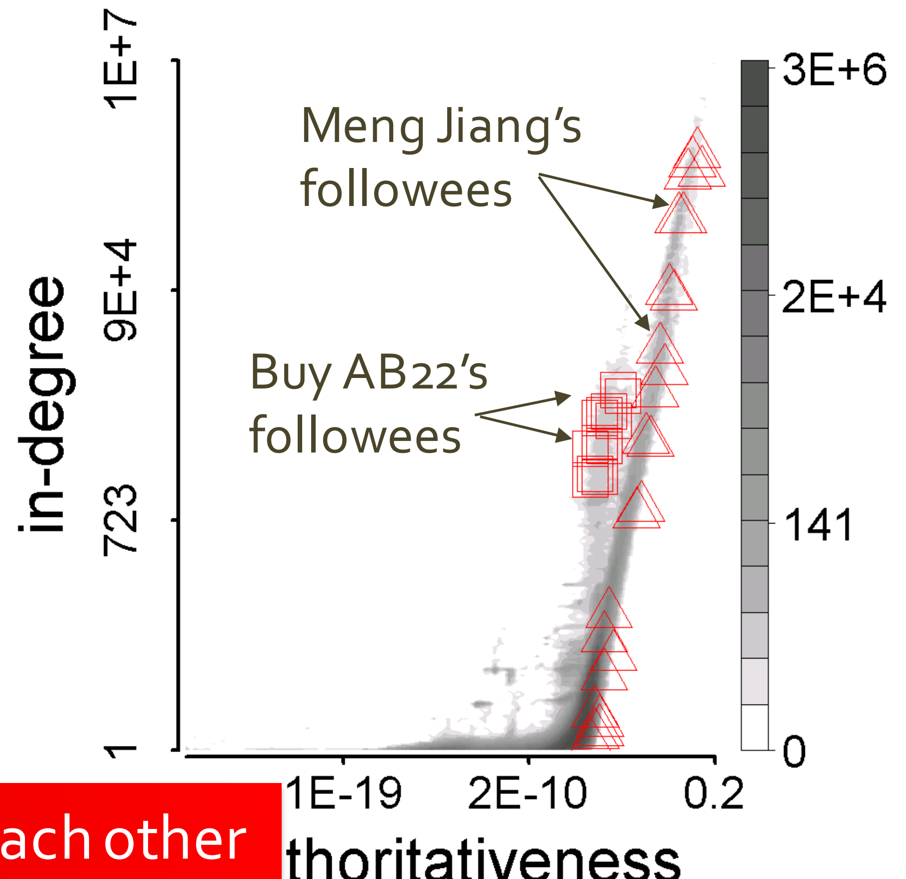
Jiang, et al. (KDD 2014)

How We/They Connect to Our/Their Followees



Jiang, et al. (KDD 2014)

How We/They Connect to Our/Their Followees

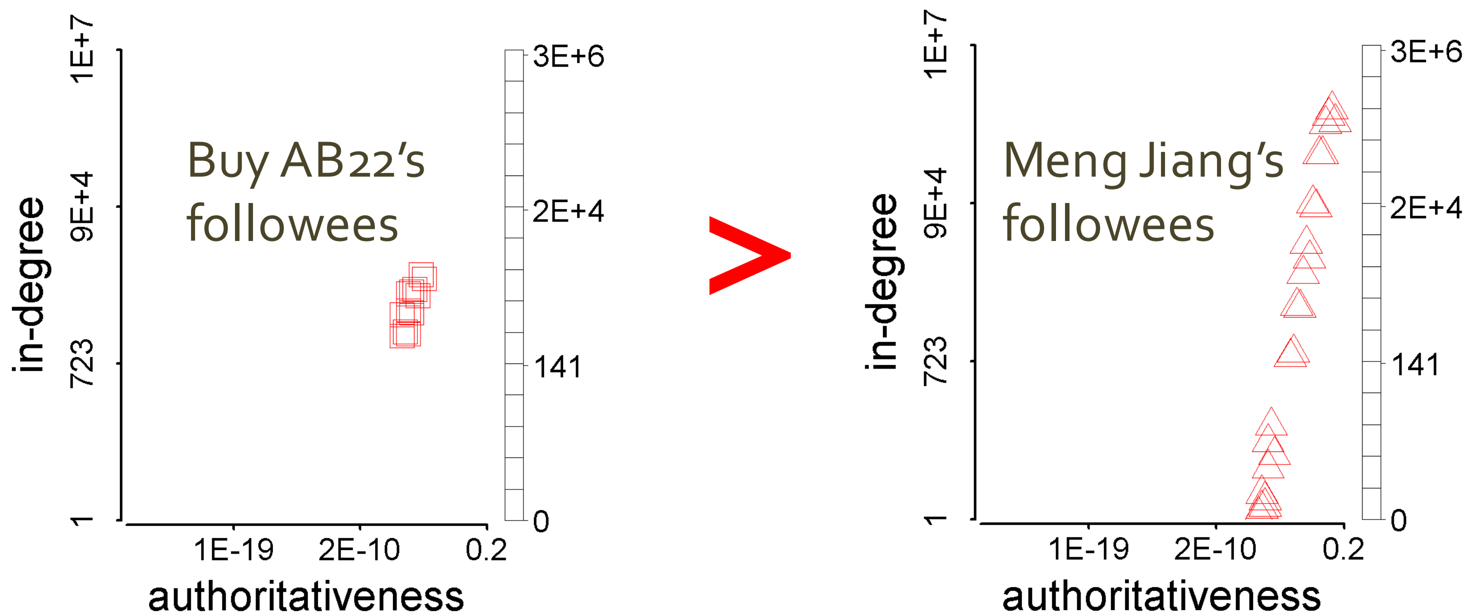


Synchronized: too similar with each other
Abnormal: too different from the majority

Jiang, et al. (KDD 2014)

Definition: Synchronicity

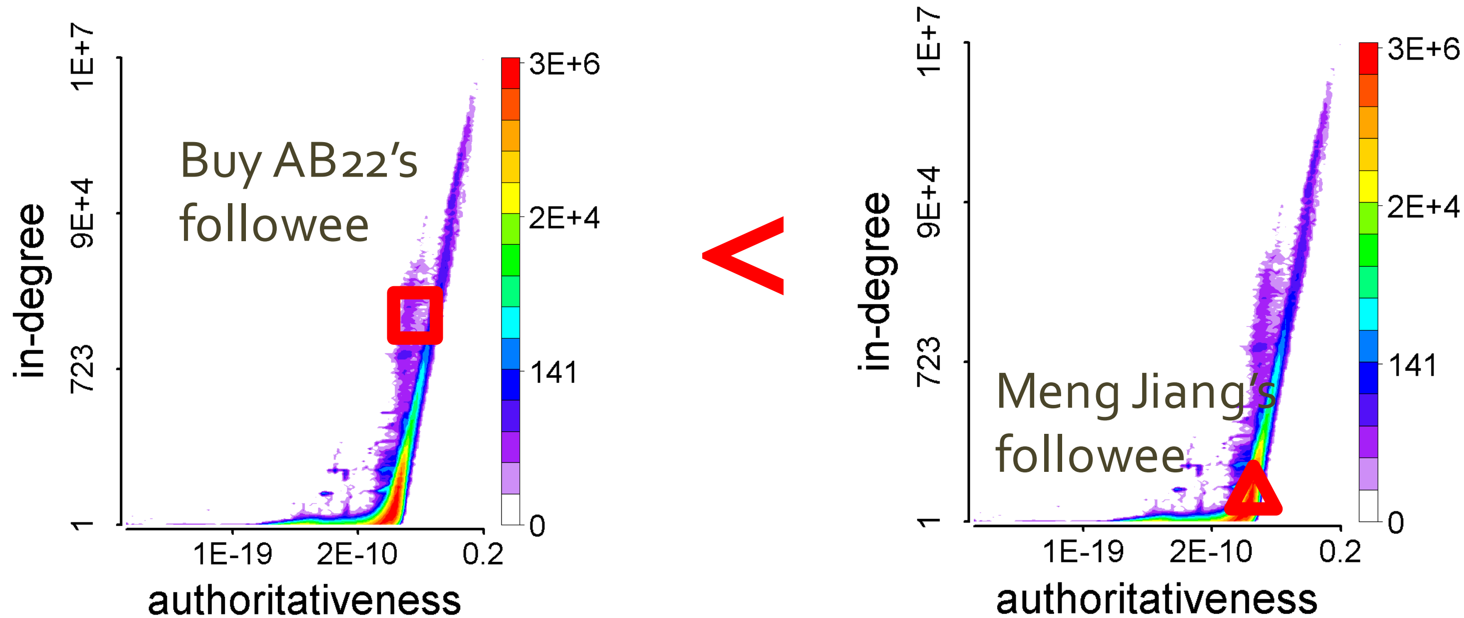
$$sync(u) = \frac{\sum_{(v,v') \in \mathcal{F}(u) \times \mathcal{F}(u)} \mathbf{p}(v) \cdot \mathbf{p}(v')}{d(u) \times d(u)}$$



Jiang, et al. (KDD 2014)

Definition: 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. (KDD 2014)

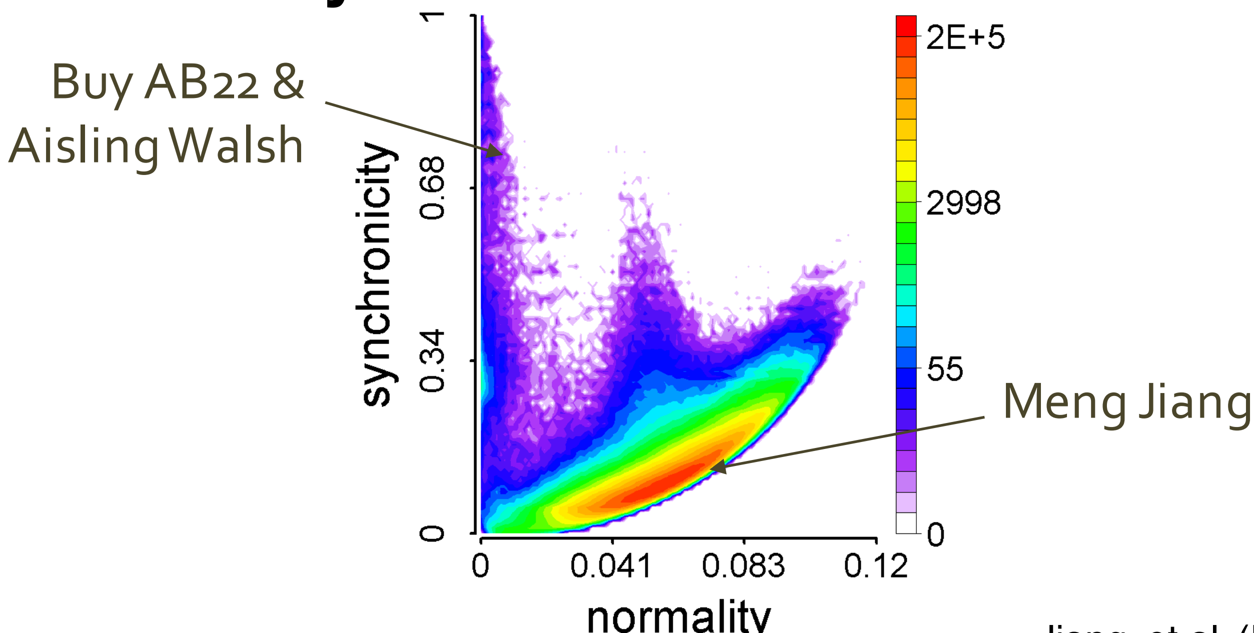
When is the Synchronicity Too High?

Problem: Given a normality value (n) of a follower, find the minimal synchronicity value (s_{min}).

Theorem:

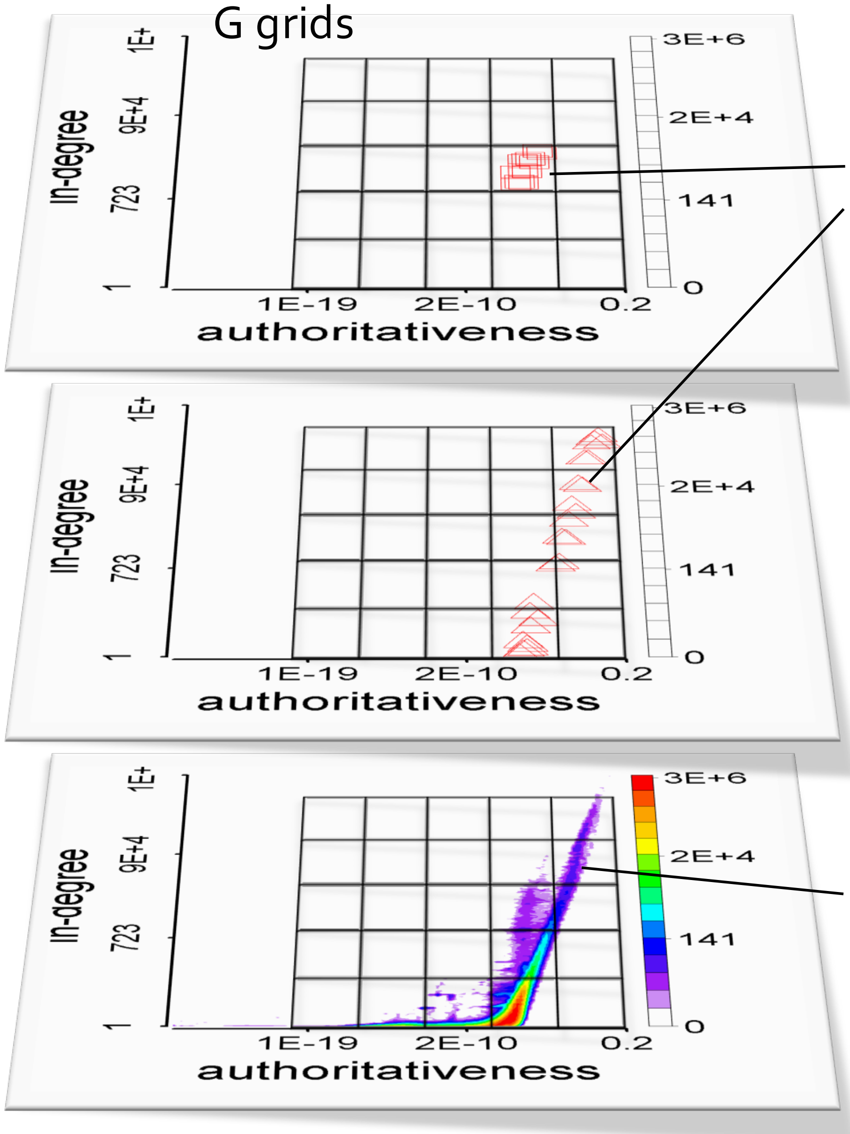
$$s_{min} = \frac{-G\,n^2 + 2\,n - s_b}{1 - G\,s_b} \qquad \textit{(parabolic lower limit)}$$

Our CatchSync:



Jiang, et al. (KDD 2014)

Proof



fp_g: #foreground points in grid g
$\sum fp_g = F = d(u)$ (#followees of u)

Given normality

$$n = \sum (fp_g/F)\,(bp_g/B) = \sum f_g b_g,$$

find minimal synchronicity

$$s_{min} = \sum (fp_g/F)\,(fp_g/F) = \sum f_g^{\,2}$$

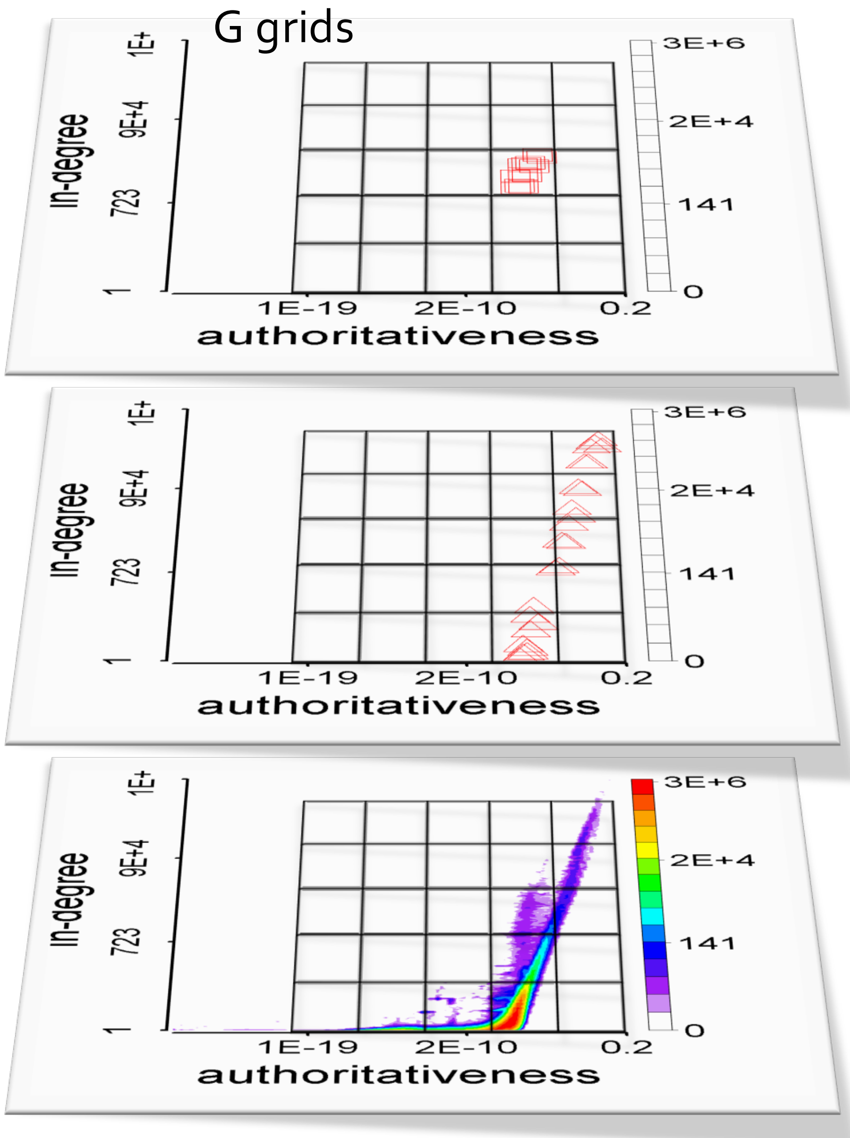
where

$$\sum f_g = 1, \sum b_g = 1$$

bp_g: #background points in grid g
$\sum bp_g = B = N$ (#all users)

Jiang, et al. (KDD 2014)

Proof



Lagrange multiplier:

minimize $s(f_g) = \sum f_g^2$

subject to $\sum f_g = 1$, $\sum f_g b_g = n$

Lagrange function:

$F(f_g, \lambda, \mu) = (\sum f_g^2) + \lambda (\sum f_g - 1) + \mu (\sum f_g b_g - n)$

Gradients:

$$\begin{cases} \nabla f_g F = 2 f_g + \lambda + \mu b_g = 0 \\ \nabla \lambda F = \sum f_g - 1 = 0 \\ \nabla \mu F = \sum f_g b_g - n = 0 \end{cases}$$

$$\begin{cases} 2 + \lambda G + \mu = 0 & \xleftarrow{\sum} \\ 2 n + \lambda + \mu s_b = 0 & \xleftarrow{\times b_g \sum} \\ 2 s_{min} + \lambda + \mu n = 0 & \xleftarrow{\times f_g \sum} \end{cases}$$

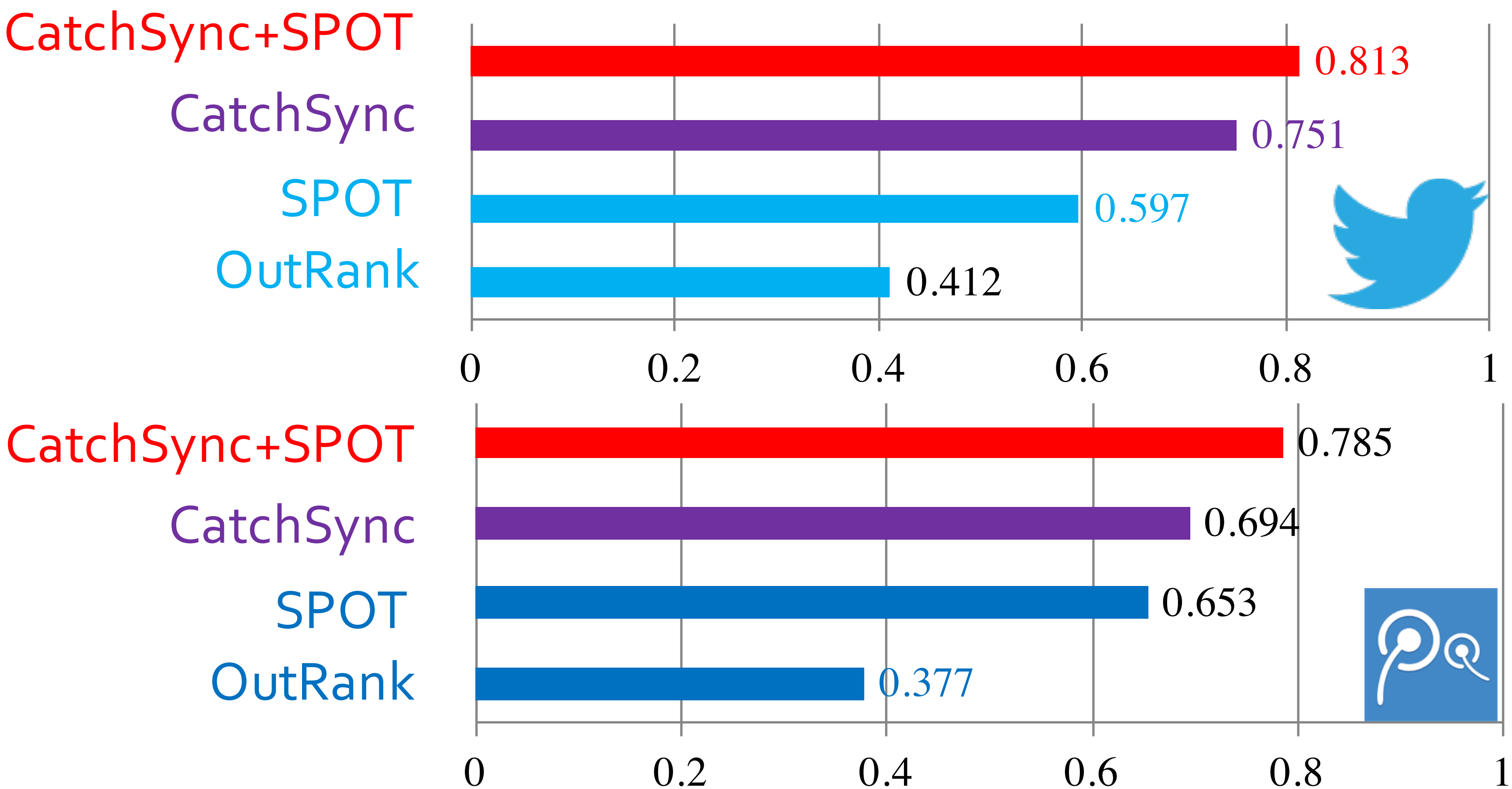
where $s_b = \sum b_g^2$.

Therefore,

$$s_{min} = \frac{-G n^2 + 2 n - s_b}{1 - G s_b}$$

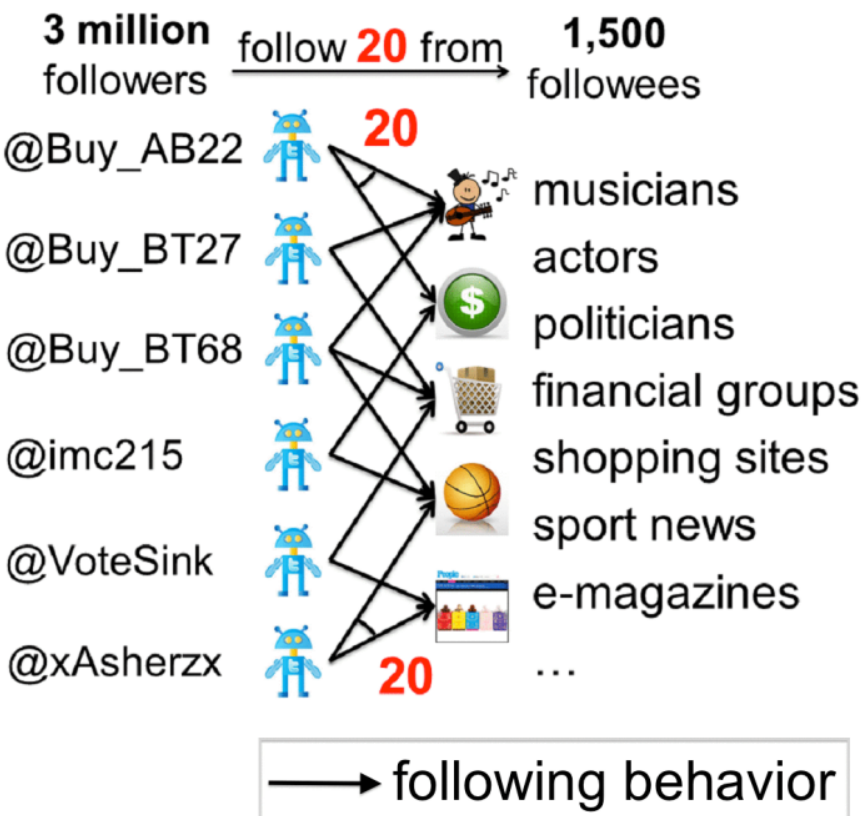
Jiang, et al. (KDD 2014)

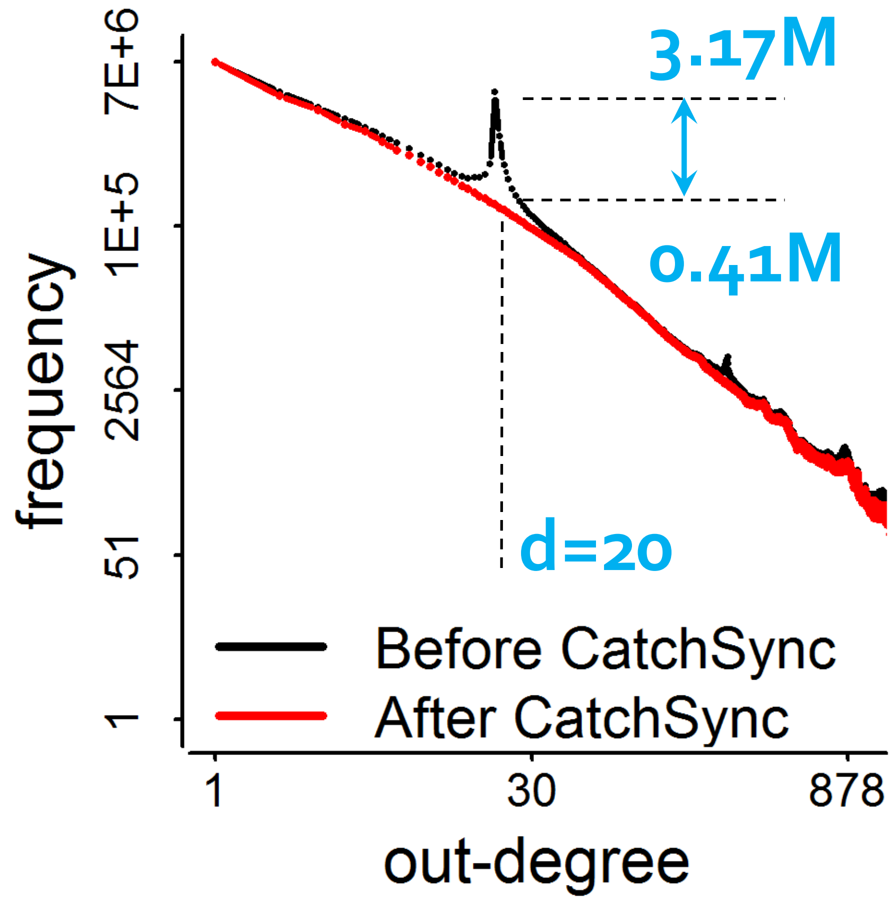
Accuracy: Complementary with Content-based Methods (SPOT)



Jiang, et al. (KDD 2014)

The Distribution was Recovered!





Jiang, et al. (KDD 2014)

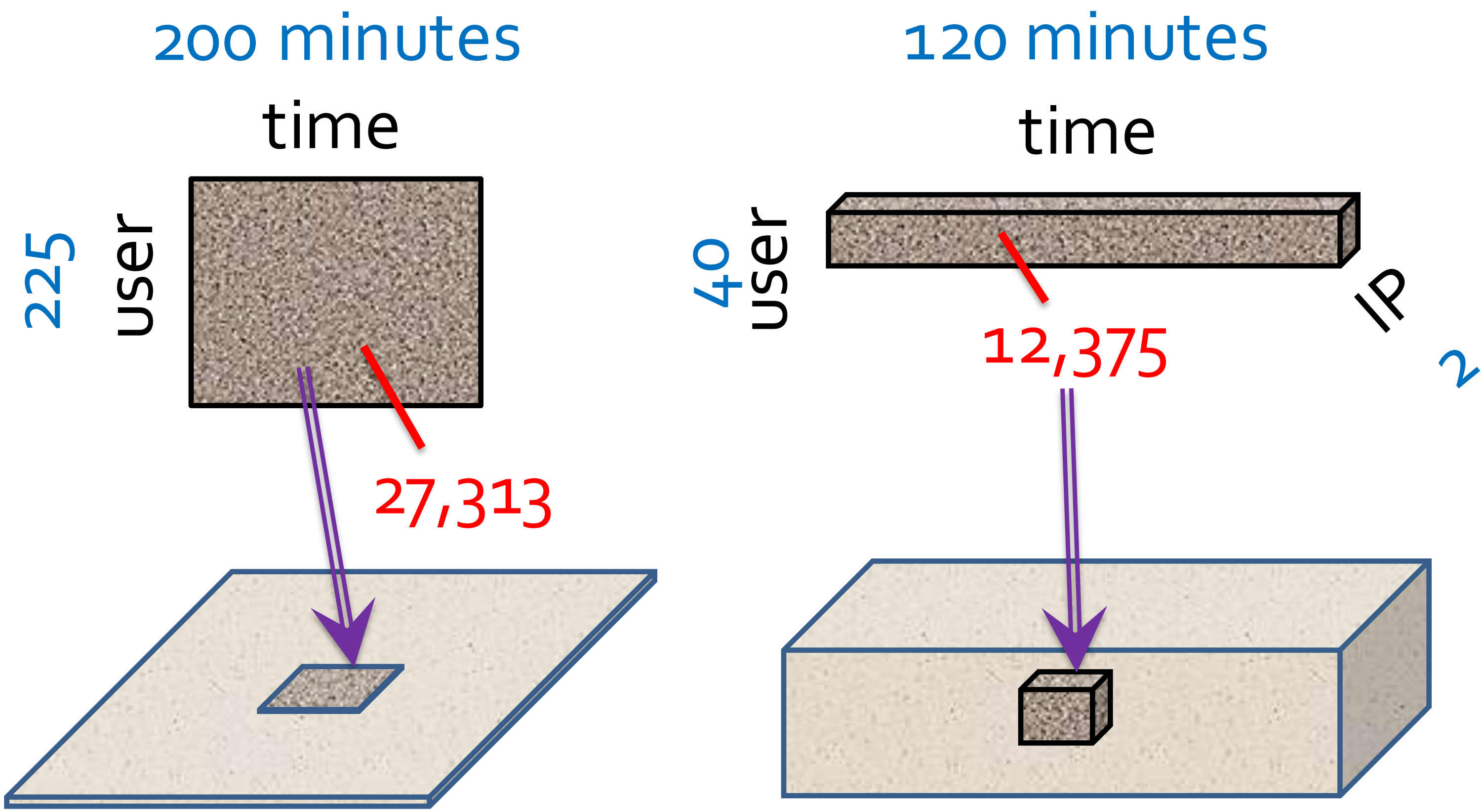
3. Social Spam

- Problem definition: Given **multidimensional behavioral data** of **spatiotemporal contexts**, find suspicious behaviors.

Dataset	Dimension/Mode				Mass
Weibo's Retweeting	User	Root ID	IP	Time (min)	#retweet
	29.5M	19.8M	27.8M	56.9K	211.7M
Weibo's 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

Meng Jiang, Alex Beutel, Peng Cui, Bryan Hooi, Shiqiang Yang, Christos Faloutsos. "Spotting Suspicious Behaviors in Multimodal Data: A General Metric and Algorithms", TKDE 2016.

Suspiciousness in Multi-dimensional Data



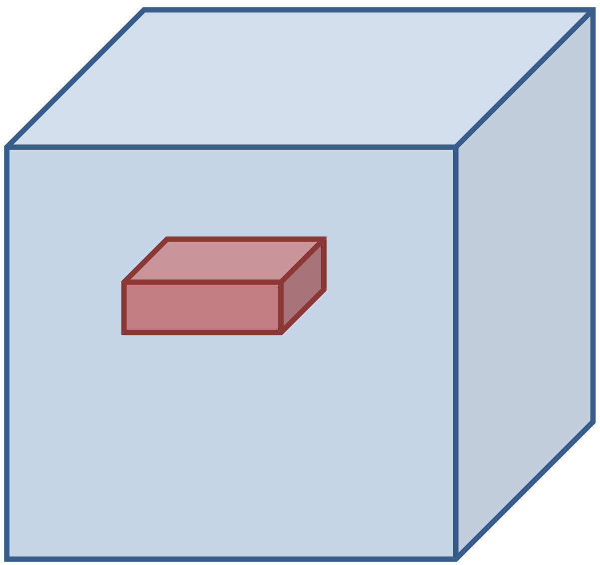
Q: Which is more suspicious?

We need a metric to evaluate the suspiciousness.

Jiang, et al. (TKDE 2016)

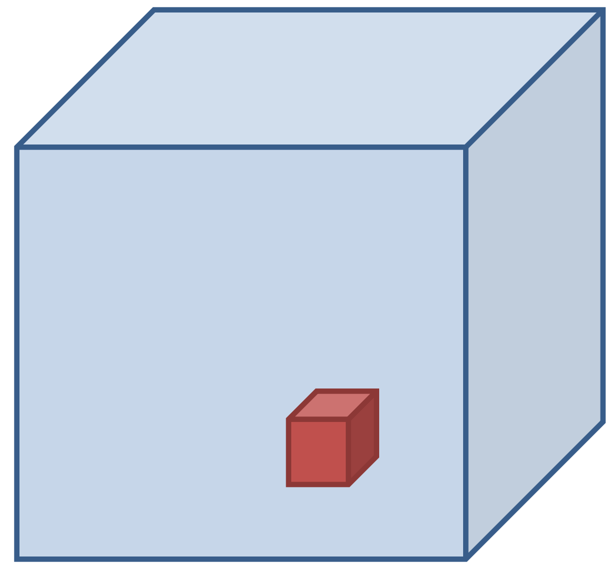
Criteria for Suspiciousness Metric

What properties are required of a good metric?



$N_1 \times N_2 \times N_3$

Count data with total "mass" C

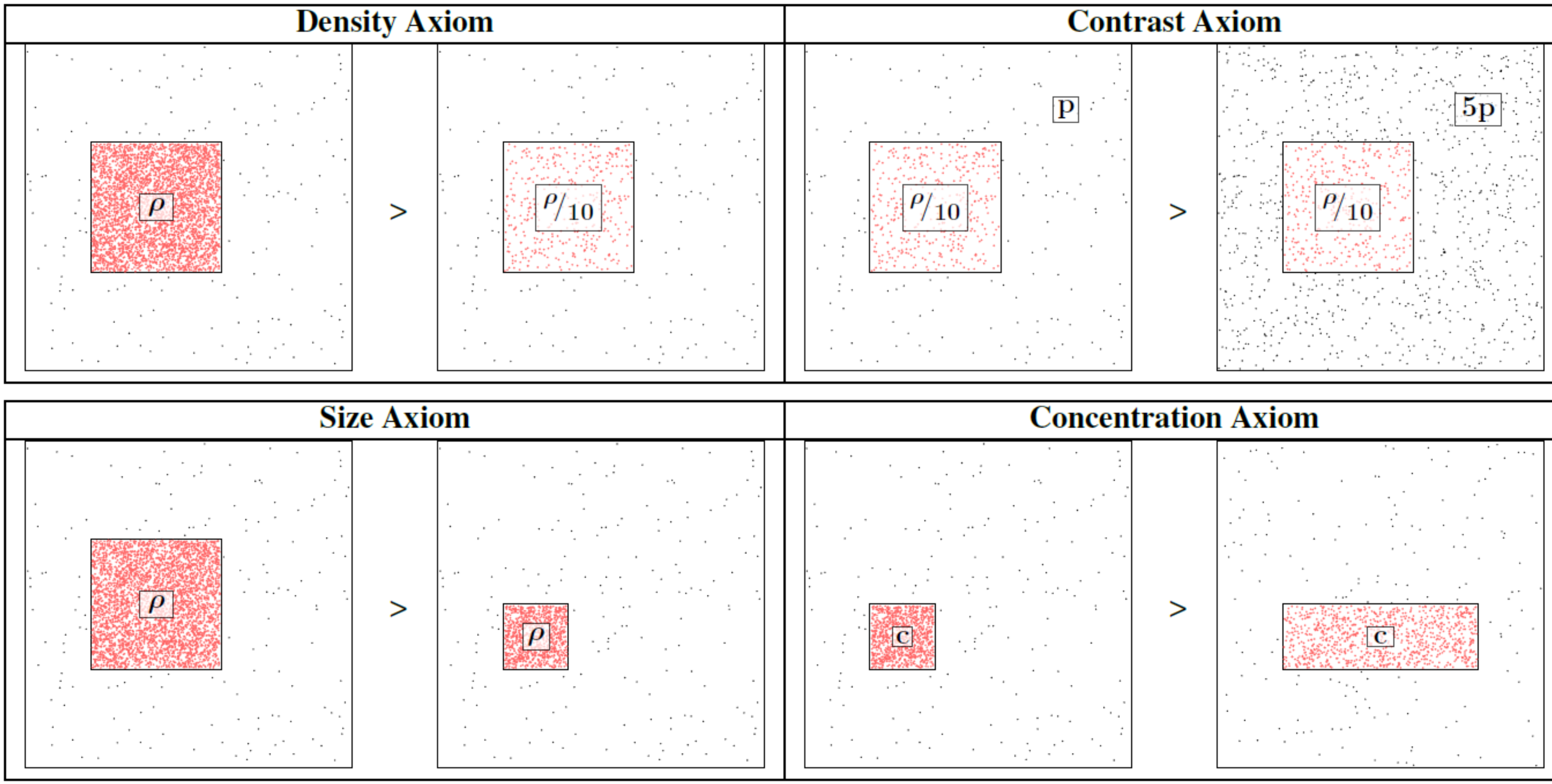


$f\left(\begin{matrix} n_1 \times n_2 \times n_3 \\ \text{mass } c \\ \text{density } \rho \end{matrix} \right)$ vs $f\left(\begin{matrix} n'_1 \times n'_2 \times n'_3 \\ \text{mass } c' \\ \text{density } \rho' \end{matrix} \right)$

Jiang, et al. (TKDE 2016)

Axioms: 1 to 4

$$c_1 > c_2 \iff f(\mathbf{n}, c_1, \mathbf{N}, C) > f(\mathbf{n}, c_2, \mathbf{N}, C)$$



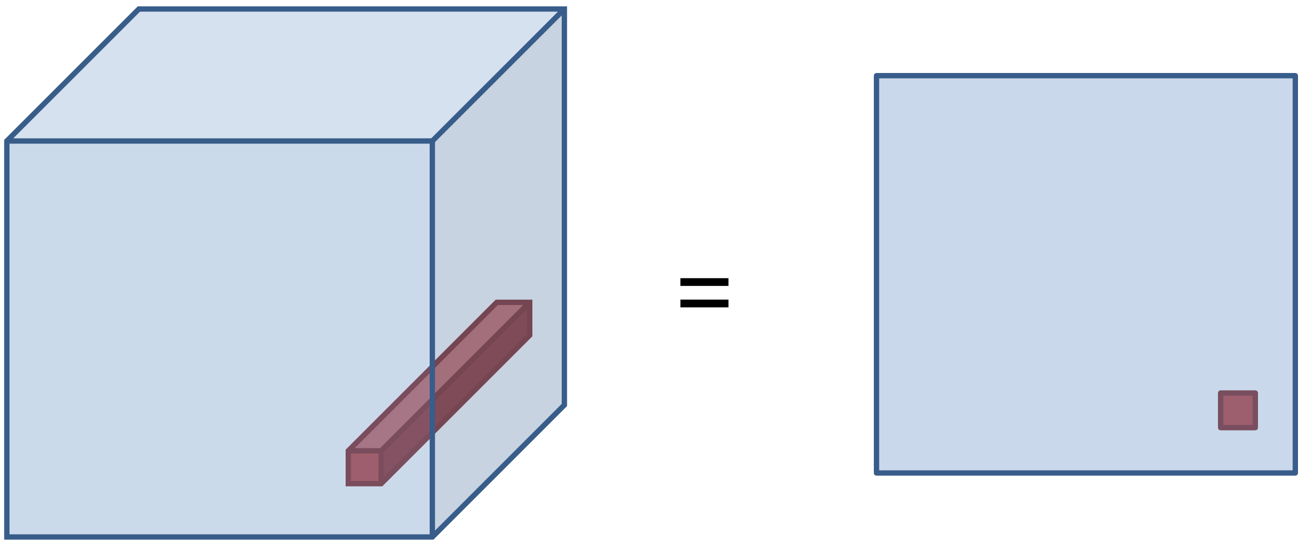
$$p_1 < p_2 \iff \hat{f}(\mathbf{n}, \rho, \mathbf{N}, p_1) > \hat{f}(\mathbf{n}, \rho, \mathbf{N}, p_2)$$

Jiang, et al. (TKDE 2016)

Axiom 5: Cross Dimensions

$$f_{K-1}\left([n_k]_{k=1}^{K-1}, c, [N_k]_{k=1}^{K-1}, C\right) = f_K\left(([n_k]_{k=1}^{K-1}, N_K), c, [N_k]_{k=1}^{K}, C\right)$$

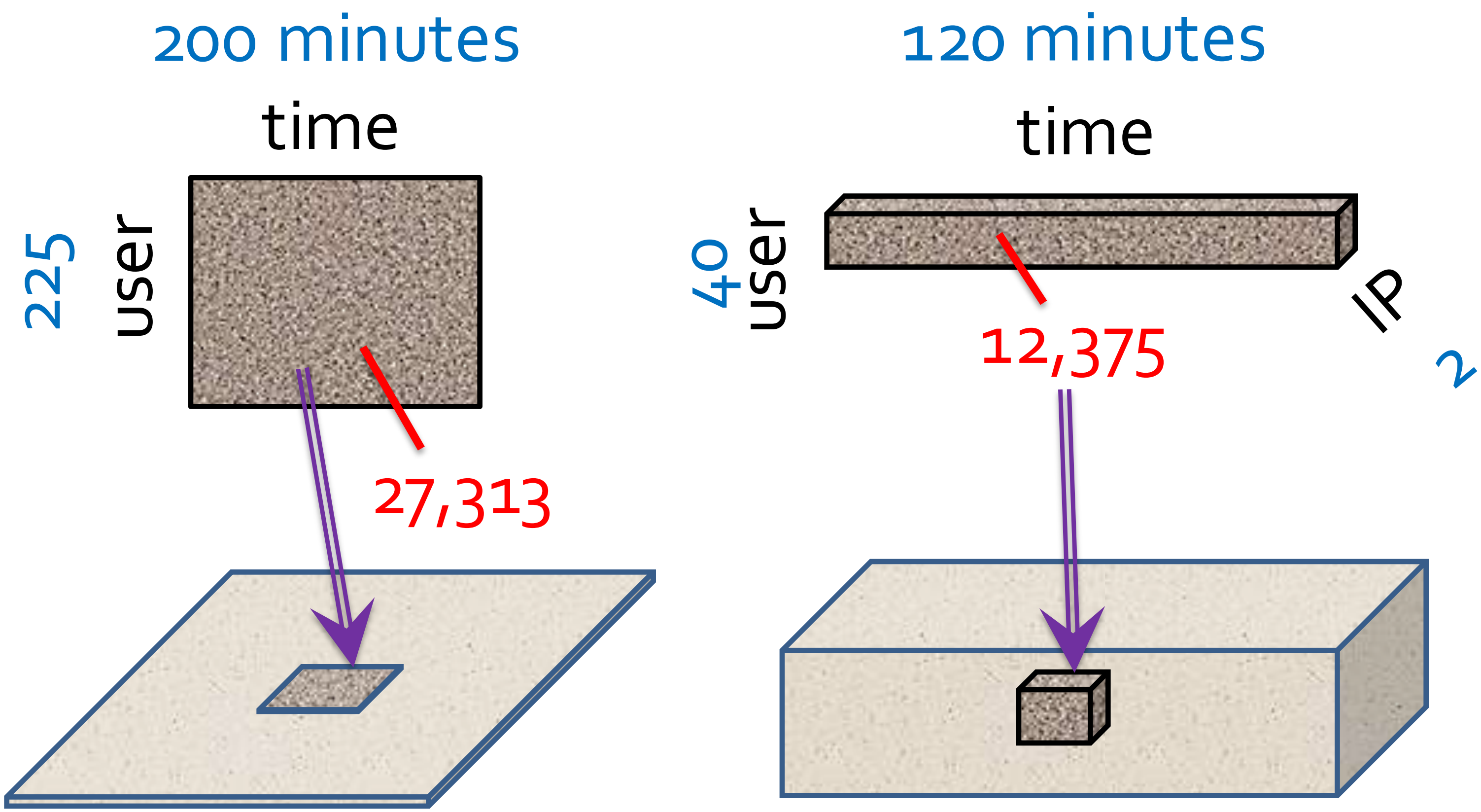
Not including a mode is the same as including all values for that mode.



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

Jiang, et al. (TKDE 2016)

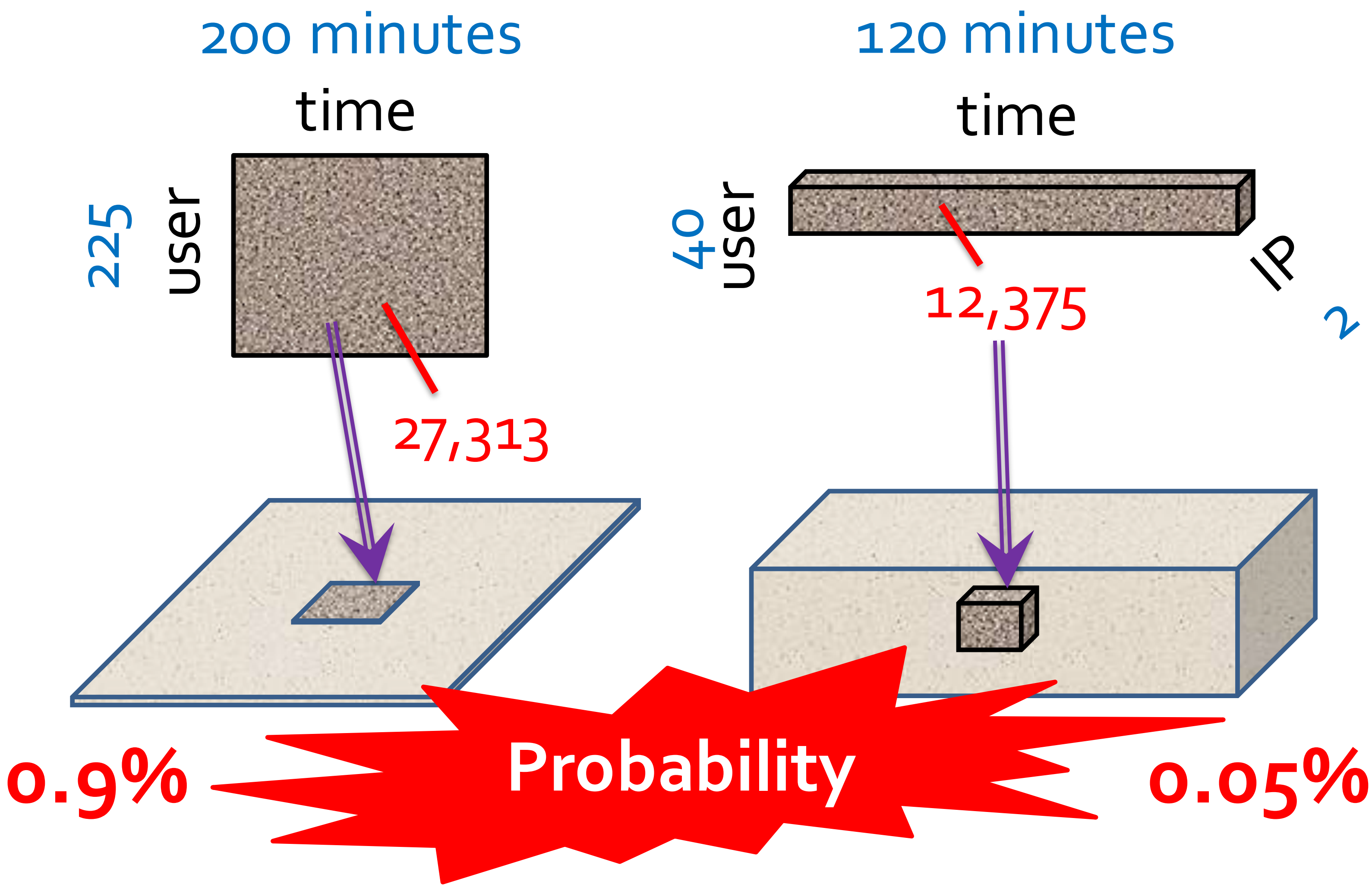
Scoring Suspiciousness



Q: Which is more suspicious?

Jiang, et al. (TKDE 2016)

Scoring Suspiciousness



Jiang, et al. (TKDE 2016)

A General Suspiciousness Metric

- 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*

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

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

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

Jiang, et al. (TKDE 2016)

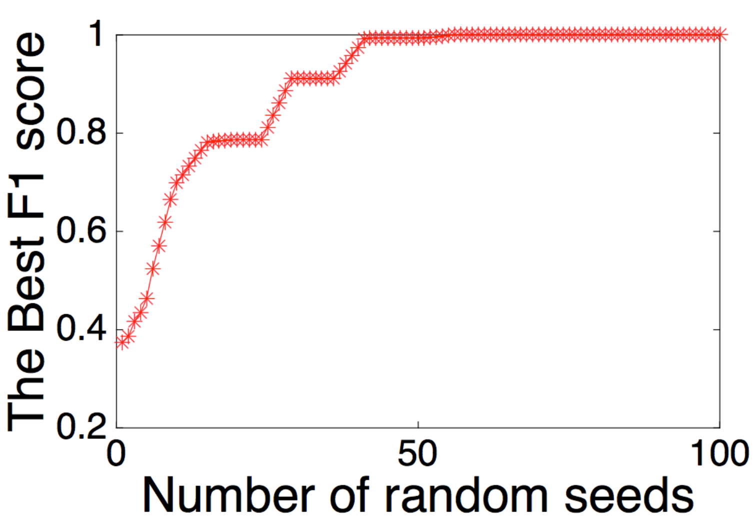
Advantages

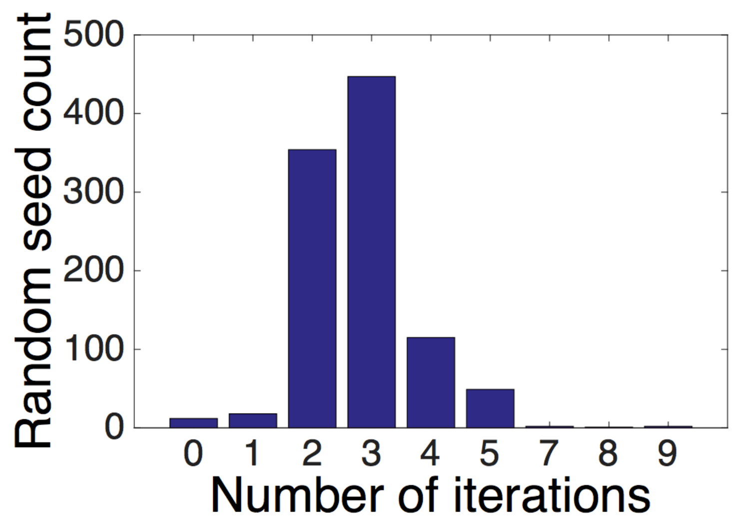
	Method	Scores Blocks	Axioms: Density 1	Size 2	Concentration 3	Contrast 4	Multi-modal 5
Metrics	**SUSPICIOUSNESS**	✔	✔	✔	✔	✔	✔
	Mass	✓	✓	✗	✗	✗	✓
	Density	✓	✓	✗	✓	✗	✗
	Average Degree [9]	✓	✓	✗	✗	✗	N/A
	Singular Value [10]	✓	✓	✓	✓	✗	✗
Methods	**CROSSSPOT**	✔	✔	✔	✔	✔	✔
	Subgraph [30, 10, 36]	✓	✓	✓	✓	✗	N/A
	CopyCatch [6]	✓	✓	✓	✓	✗	N/A
	EigenSpokes [31]	✗	N/A				
	TrustRank [14, 8]	✗	N/A				
	BP [28, 1]	✗	N/A				

Jiang, et al. (TKDE 2016)

CrossSpot Algorithm

- Greedy algorithm by maximizing the metric
  - Start with seed blocks
  - Parameter-free: iteratively update the blocks
  - Scalable: parallelize to multiple machines





Jiang, et al. (TKDE 2016)

Hijacked Hashtags

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. (TKDE 2016)

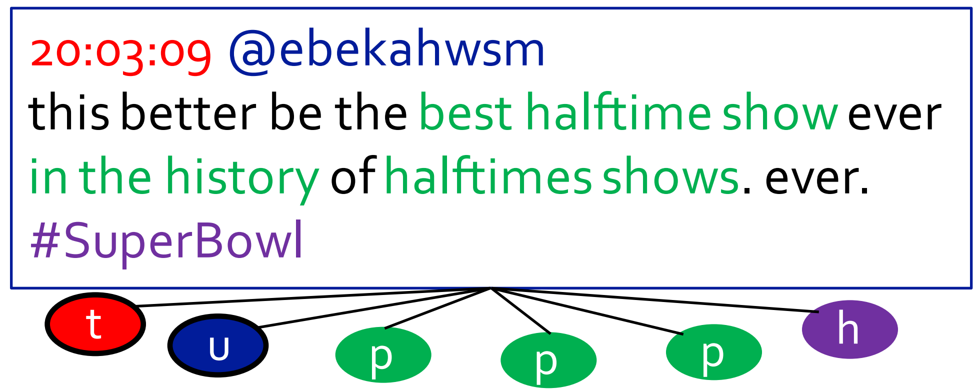
Network Attacks (LBNL)

	#	Src-IP×dst-IP×port×second	Mass c	Suspiciousness
CrossSpot	1	411×9×6×**3,610**	47,449	552,465
	2	533×6×1×**3,610**	30,476	400,391
	3	5×5×2×**3,610**	18,881	317,529
	4	11×7×7×**3,610**	20,382	295,869
HOSVD	1	15×1×1×1,336	4,579	80,585
	2	1×2×2×1,035	1,035	18,308
	3	1×1×1×1,825	1,825	34,812
	4	1×13×6×181	1,722	29,224

Jiang, et al. (TKDE 2016)

4. Events or Advertising Campaigns

- Density in **multi-contextual** behavioral data



Contextual factors:

Dynamic

One-guaranteed value

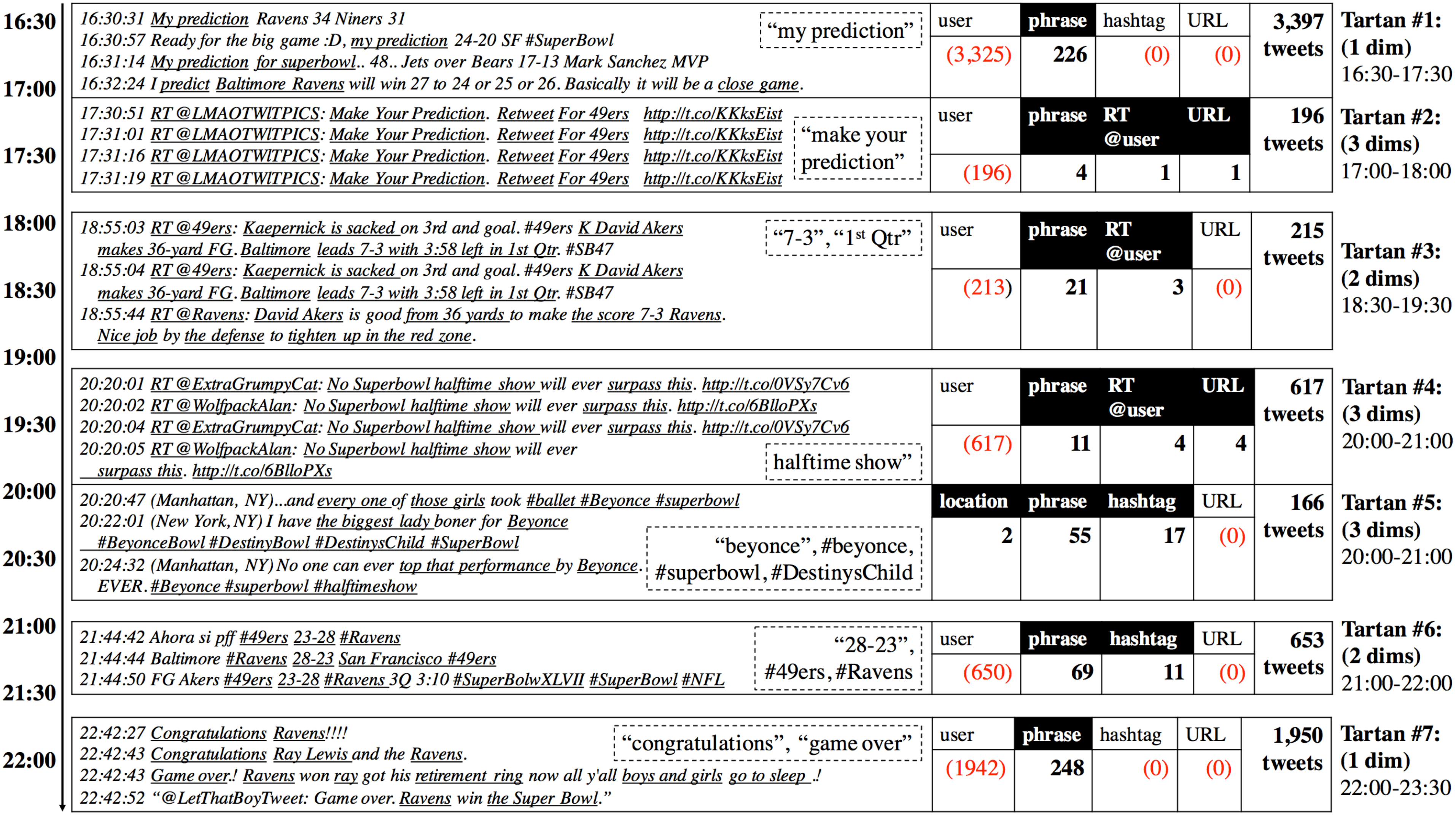
Empty (set of) *value*

Set value

Empty (set of) *value*

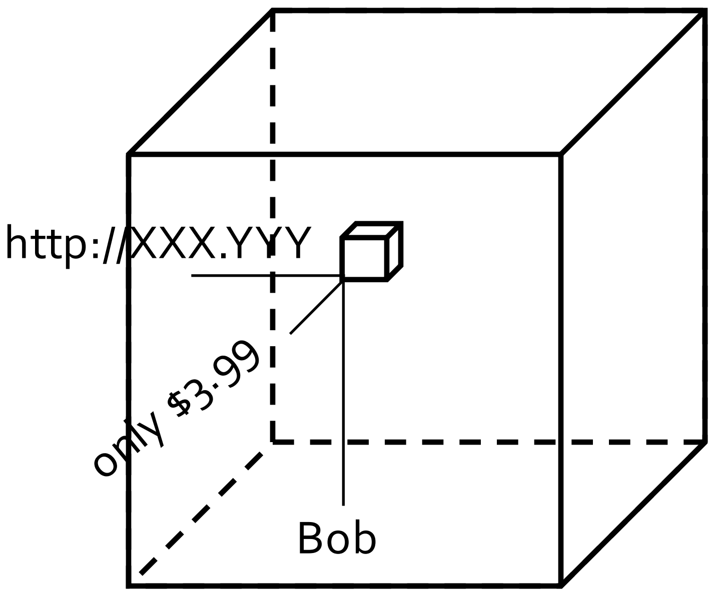
Time slice	User	Location	Phrase	Hashtag	URL
20:00-20:30	@ebekahwsm	∅	{best halftime show, in the history, halftimes shows}	{#SuperBowl}	∅

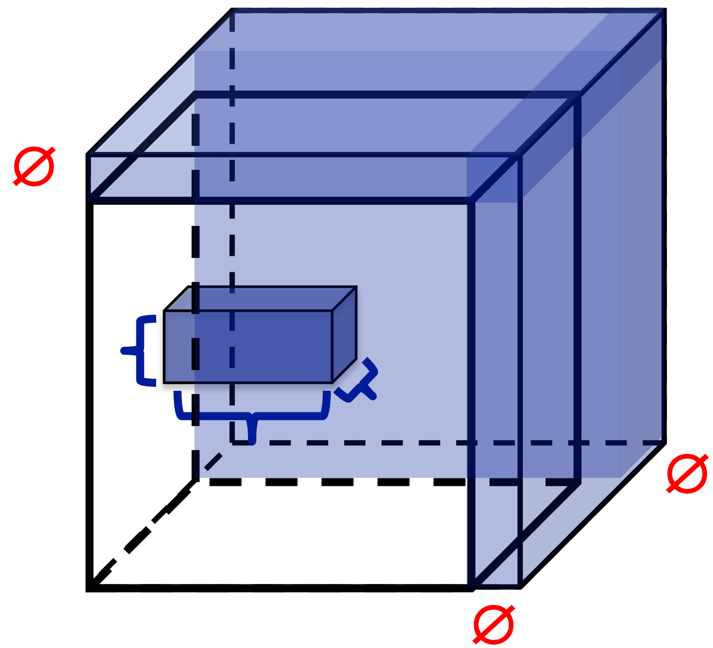
Given tweets in Super Bowl 2013, can we find events (score prediction, halftime show, advertising campaigns, etc.)?



Meng Jiang, Christos Faloutsos, Jiawei Han. “CATCHTARTAN: Representing and Summarizing Dynamic Multicontextual Behaviors”, KDD 2016.

Tensor Fails: Representation



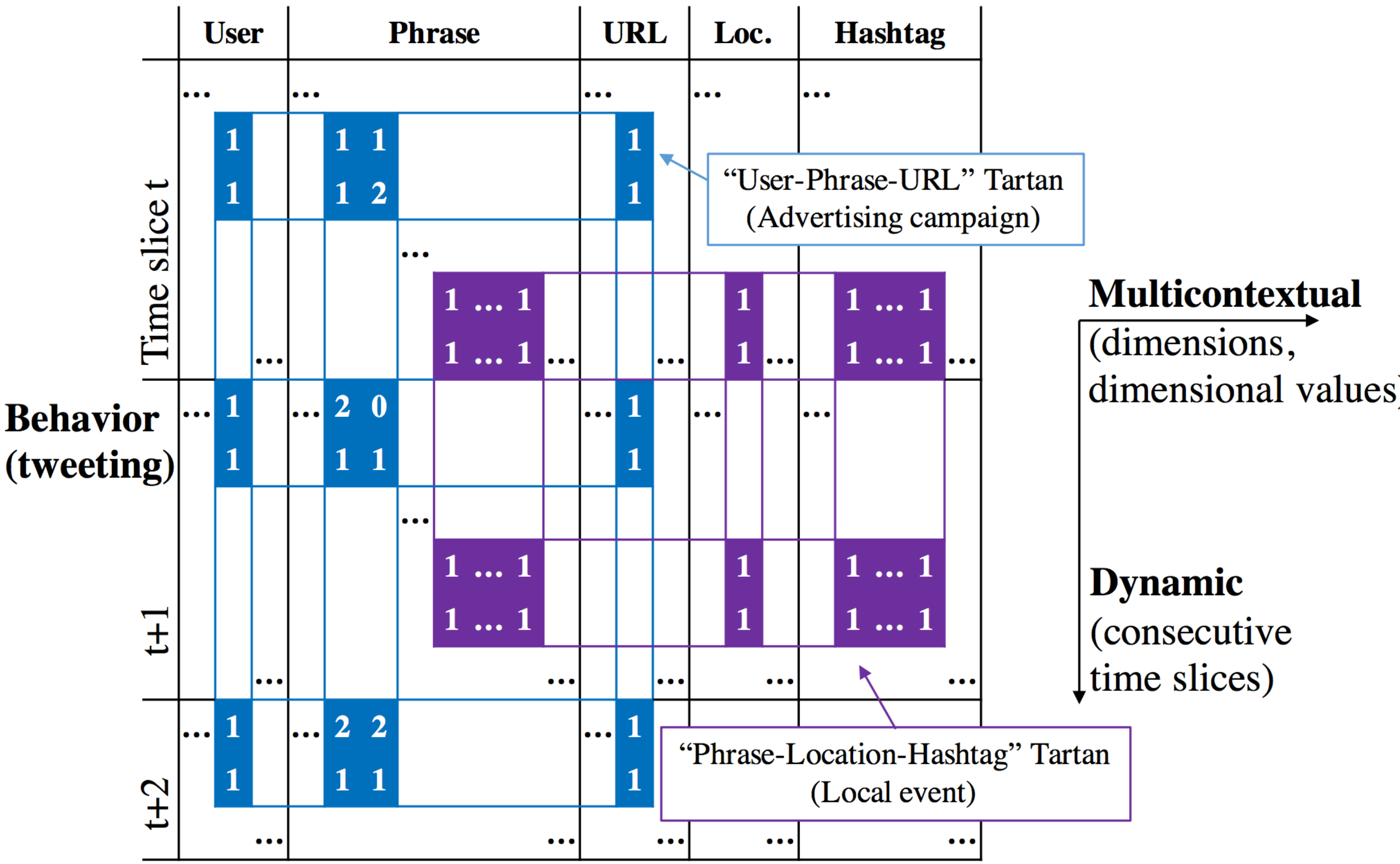


Jiang, et al. (KDD 2016)

"Two-Level Matrix" and Tartans

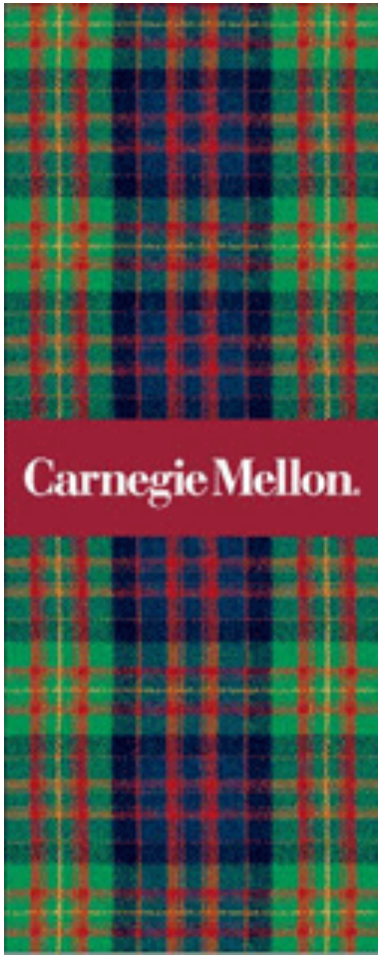
Behavior representation

Behavior summaries



Jiang, et al. (KDD 2016)

What is Tartan?



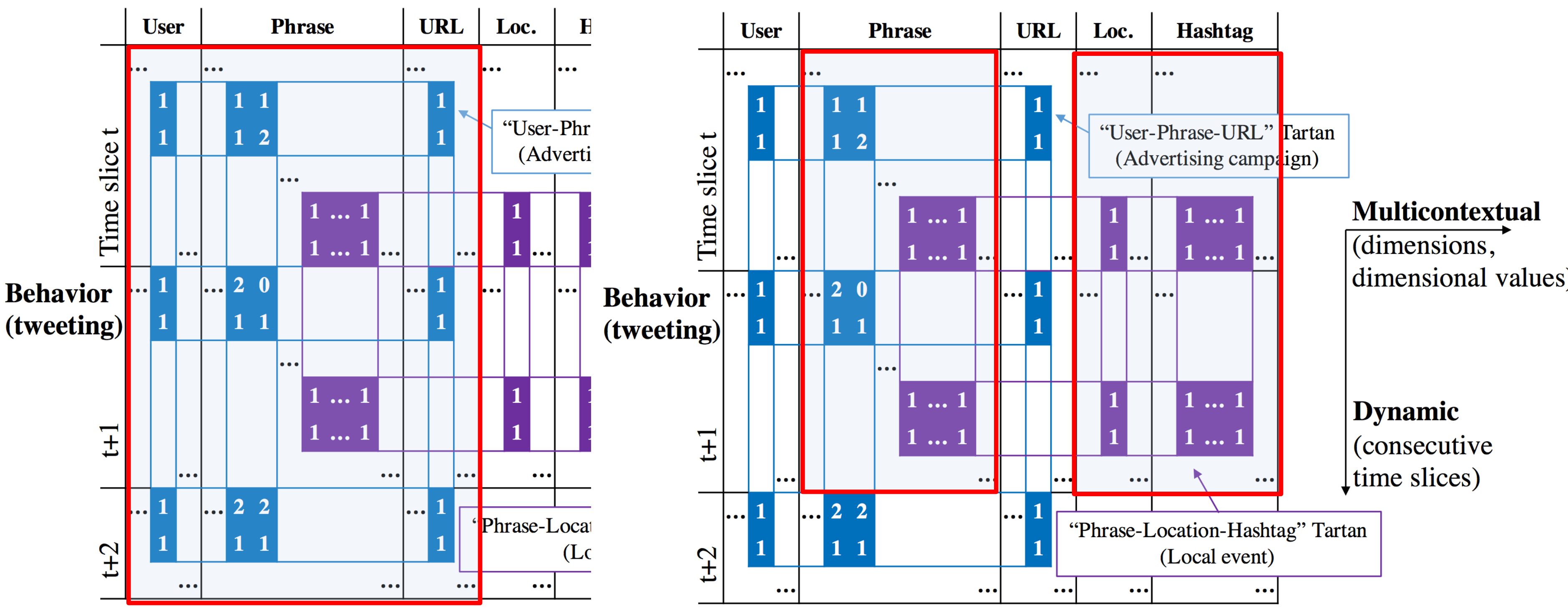
Visited CMU in 2012-13

Watched lots of Tartans' games…

Objective Function to Maximize (Minimum Description Length principle)

$$\boxed{f(\mathcal{A}, \mathcal{X}) = L(\mathcal{X}^{\mathcal{A}}) - L(\mathcal{A}) - L(\mathcal{X}^{\mathcal{A}} \backslash \mathcal{A}).}$$

Tartan **Data** **First-level matrix** **Individual entries**



$$\mathcal{X}^{\mathcal{A}} = \{\mathcal{X}_d^{(t)}(b,i) | d \in \mathcal{D}, t \in \mathcal{T}, i \in \{1,\ldots,N_d\}, b \in \{1,\ldots,E^{(t)}\}\}.$$

Objective Function to Maximize (cont.)

$$\boxed{f(\mathcal{A}, \mathcal{X}) = L(\mathcal{X}^{\mathcal{A}}) - L(\mathcal{A}) - L(\mathcal{X}^{\mathcal{A}} \backslash \mathcal{A}).}$$

$$\begin{aligned} V &= \left(\sum_{d \in \mathcal{D}} N_d\right)\left(\sum_{t \in \mathcal{T}} E^{(t)}\right). \\ C &= \sum_{d \in \mathcal{D}, t \in \mathcal{T}} \sum_{b \in \{1,\ldots,E^{(t)}\}, i \in \{1,\ldots,N_d\}} \mathcal{X}_d^{(t)}(b, i). \end{aligned}$$

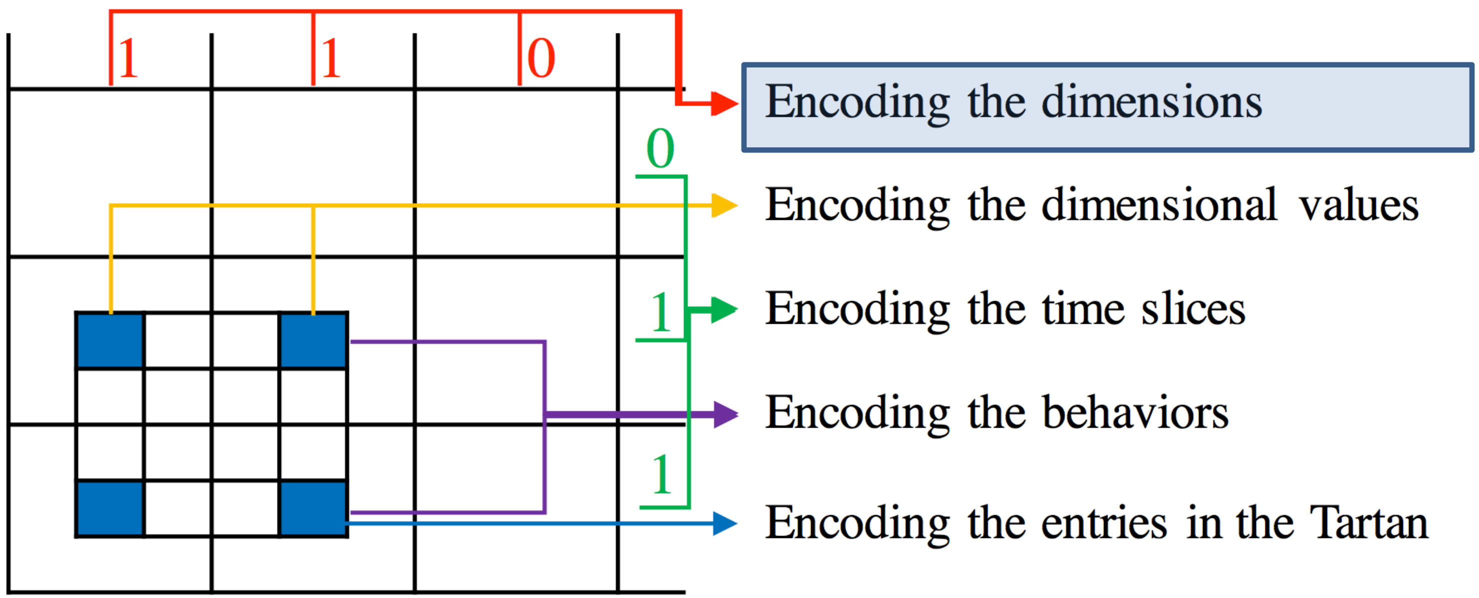
$$\boxed{\begin{aligned} L(\mathcal{X}^{\mathcal{A}}) &= g(V + C, C) + L_{\mathcal{D}}(\mathcal{A}) + L_{\mathcal{T}}(\mathcal{A}) \\ &\quad + \sum_{d \in \mathcal{D}} \log^* N_d + \sum_{t \in \mathcal{T}} \log^* E^{(t)}. \end{aligned}}$$

$$\boxed{L(\mathcal{A}) = L_{\mathcal{D}}(\mathcal{A}) + L_{\mathcal{V}}(\mathcal{A}) + L_{\mathcal{T}}(\mathcal{A}) + L_{\mathcal{B}}(\mathcal{A}) + L_{A}(\mathcal{A}).}$$

$$\boxed{L(\mathcal{X}^{\mathcal{A}} \backslash \mathcal{A}) = g(V + C - v - c, C - c);}$$

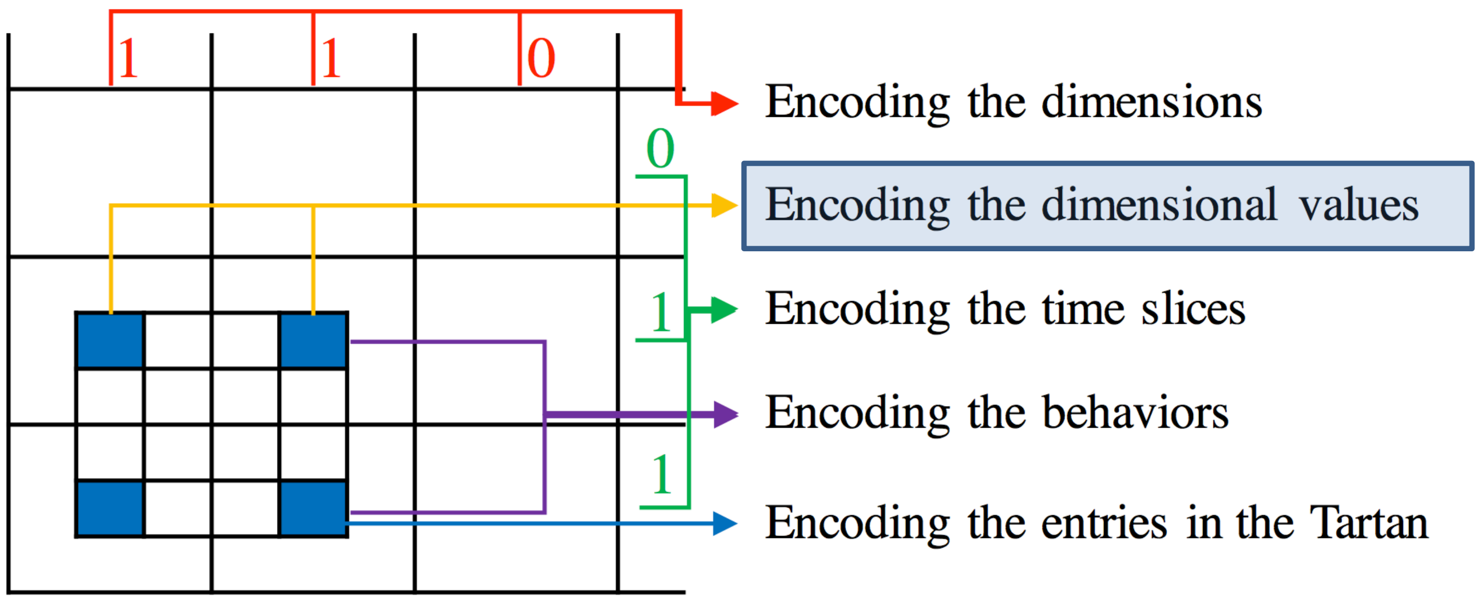
Jiang, et al. (KDD 2016)

Encoding the Tartan: Dimensions



$$
\begin{aligned}
H_{\mathcal{D}}(X) &= -\sum_{x\in\{0,1\}} P(X=x)\log P(X=x) \\
&= -\left(\frac{D^{\mathcal{A}}}{D}\log\frac{D^{\mathcal{A}}}{D} + \frac{D-D^{\mathcal{A}}}{D}\log\frac{D-D^{\mathcal{A}}}{D}\right). \\
L_{\mathcal{D}}(\mathcal{A}) &= \log^* D + \log^* D^{\mathcal{A}} + D\cdot H_{\mathcal{D}}(X) \\
&= \log^* D + \log^* D^{\mathcal{A}} + g(D, D^{\mathcal{A}}),
\end{aligned}
$$

Encoding the Tartan: Dimensional Values

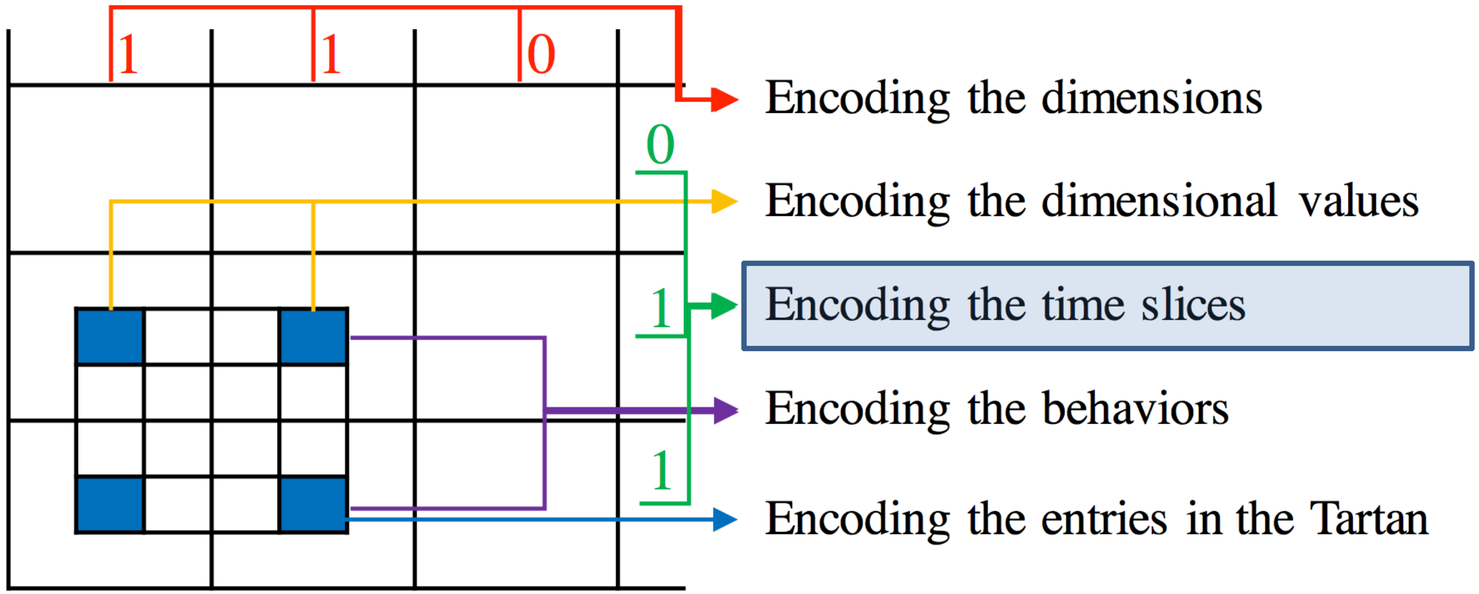


$$H_{\mathcal{V}_d}(X) \quad = \quad -\left(\frac{n_d}{N_d}\log\frac{n_d}{N_d} + \frac{N_d - n_d}{N_d}\log\frac{N_d - n_d}{N_d}\right).$$

$$L_{\mathcal{V}}(\mathcal{A}) \quad = \quad \sum_{d\in\mathcal{D}}\left(\log^* N_d + \log^* n_d + g(N_d, n_d)\right).$$

Jiang, et al. (KDD 2016)

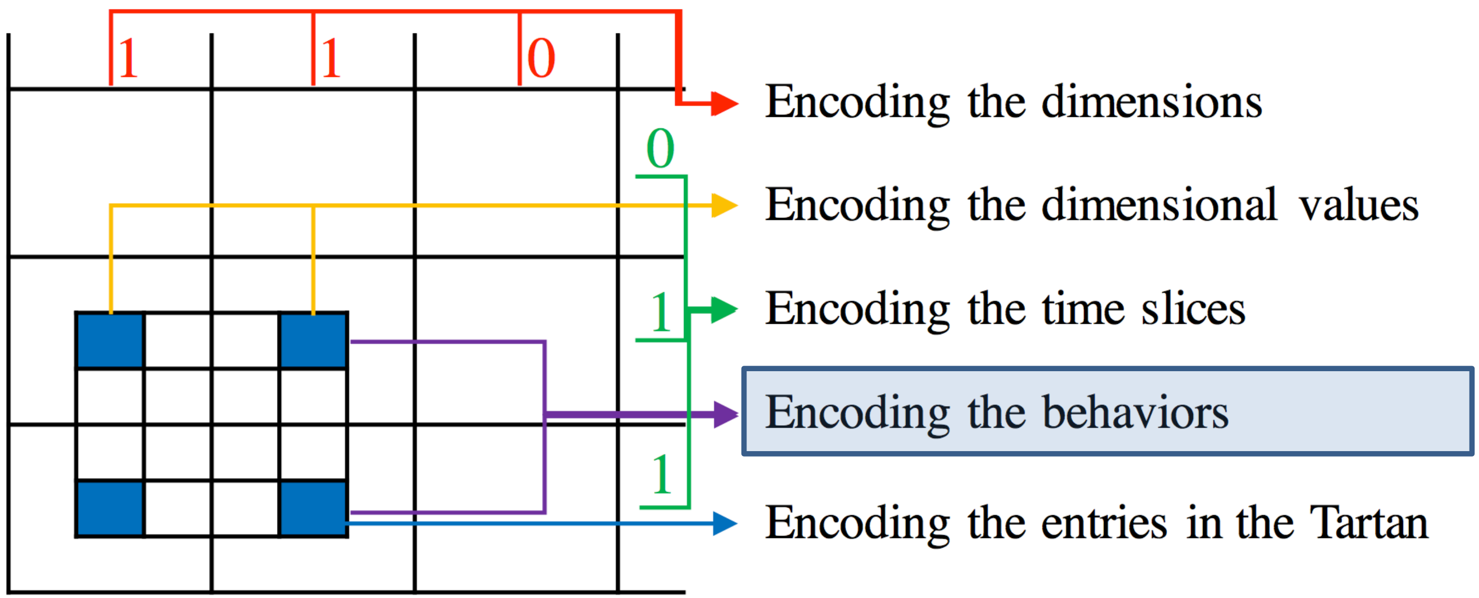
Encoding the Tartan: Time Slices



$$L_{\mathcal{T}}(\mathcal{A}) \quad = \quad \log^* T + \log^* T^{\mathcal{A}} + \log^* t_{start}$$

Jiang, et al. (KDD 2016)

Encoding the Tartan: Behaviors

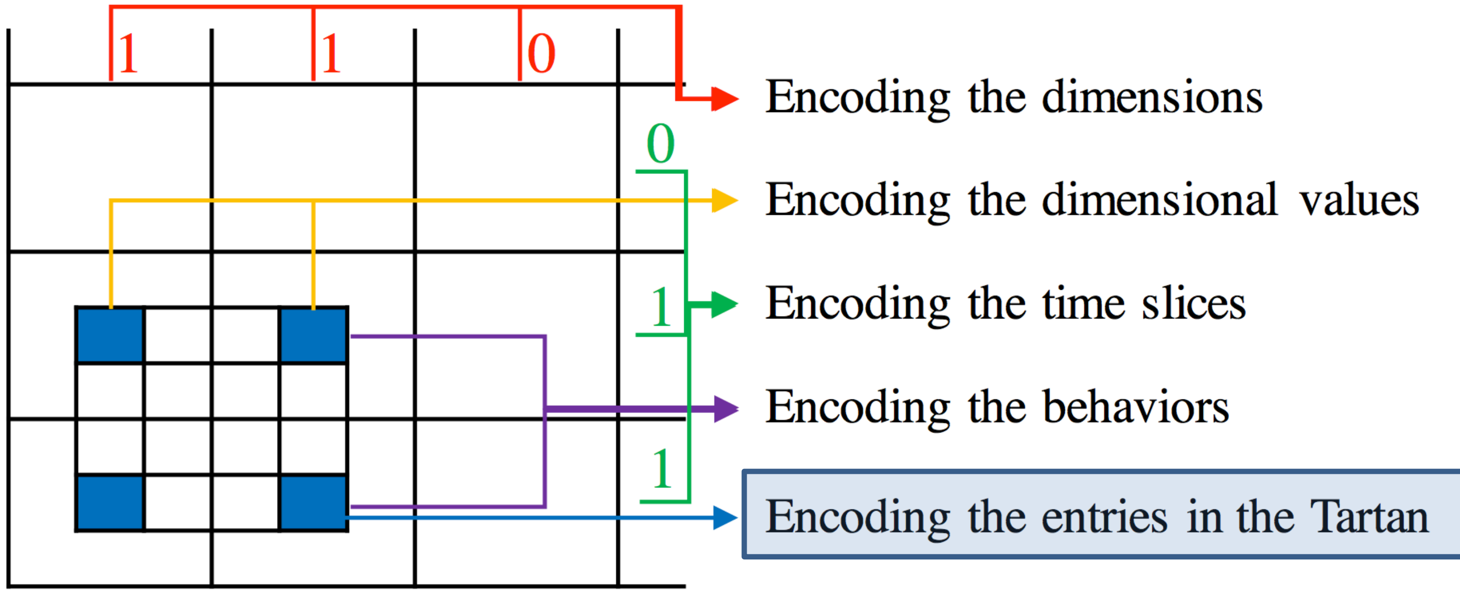


$$H_{\mathcal{B}^{(t)}}(X) = -\left(\frac{e^{(t)}}{E^{(t)}} \log \frac{e^{(t)}}{E^{(t)}} + \frac{E^{(t)}-e^{(t)}}{E^{(t)}} \log \frac{E^{(t)}-e^{(t)}}{E^{(t)}}\right).$$

$$L_{\mathcal{B}}(\mathcal{A}) = \sum_{t\in\mathcal{T}} \left(\log^* E^{(t)} + \log^* e^{(t)} + g(E^{(t)}, e^{(t)})\right).$$

Jiang, et al. (KDD 2016)

Encoding the Tartan: Entries



$$v = \left(\sum_{d\in\mathcal{D}} n_d\right)\left(\sum_{t\in\mathcal{T}} e^{(t)}\right).$$

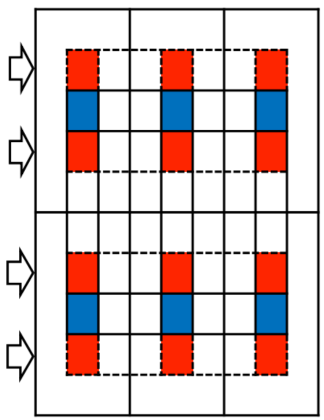
$$c = \sum_{d\in\mathcal{D},t\in\mathcal{T}} \sum_{b\in\mathcal{B}^{(t)},i\in\mathcal{V}_d} \mathcal{X}_d^{(t)}(b,i).$$

$$H_{\mathcal{A}}(X) = -\left(\tfrac{c}{v+c}\log\tfrac{c}{v+c} + \tfrac{v}{v+c}\log\tfrac{v}{v+c}\right).$$

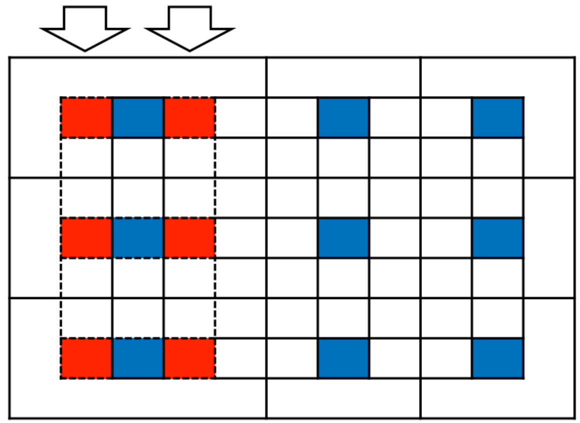
$$L_{\mathcal{A}}(\mathcal{A}) = (v+c)H_{\mathcal{A}}(X) = g(v+c,c).$$

Jiang, et al. (KDD 2016)

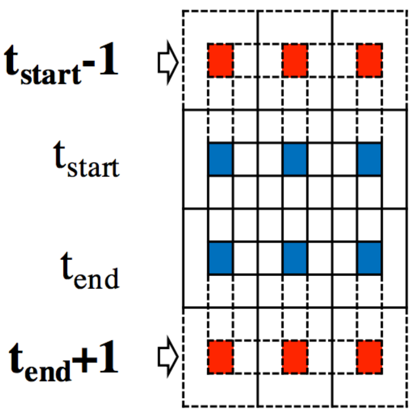
Greedy Search for the Local Optimum



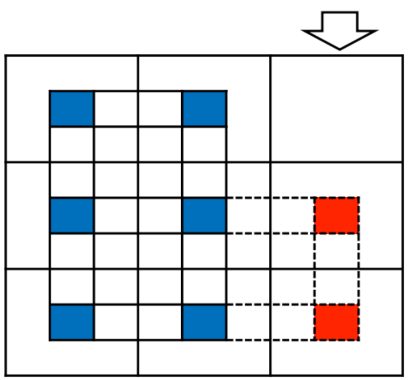
(a) Update the set of behaviors.



(b) Update the set of values.



(c) Update the consecutive time slices.



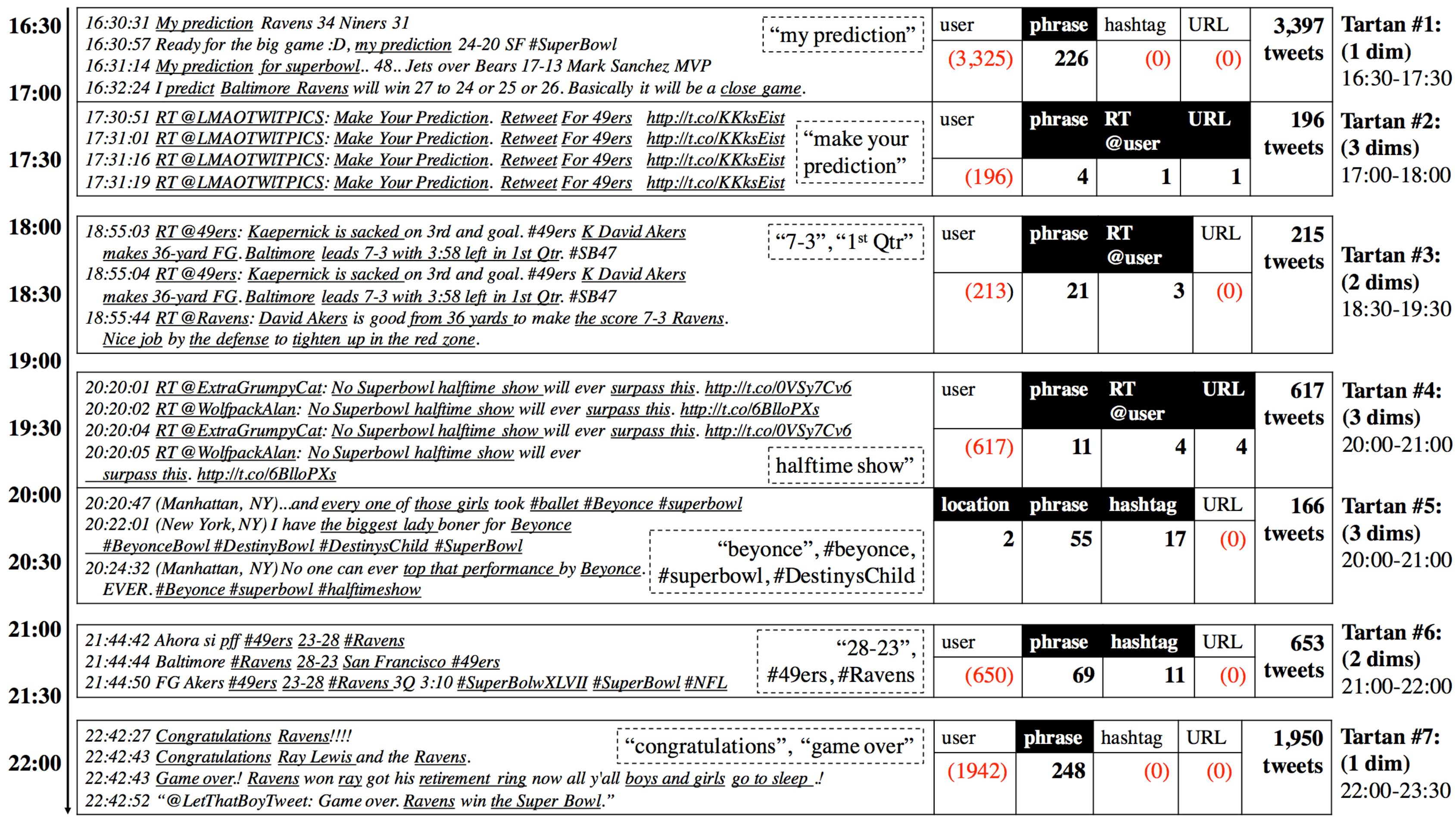
(d) Update the set of dimensions.

Time complexity:

$$\mathcal{O}(\sum_d N_d \log N_d + \sum_t E^{(t)} \log E^{(t)})$$

Jiang, et al. (KDD 2016)

Experiments: Events in Tweets



Jiang, et al. (KDD 2016)

Experiments: Research Trends in DBLP Data

1997 2000 2003 2006 2009 2012

Author	Venue	Keyword	Cited	#Paper
76 Cheng-xiang Zhai Hui Fang S. Kambhampati	**7** SIGIR VLDB TKDE	**7** "information retrieval" "data integration" "text classification"	**68** p56743[1] p62995 p76869	**32** 2003- 2007

[1] "A language modeling approach to information retrieval"

Venue	Keyword	#Paper
5 ICML NIPS …	**6** "reinforcement learning" "machine learning"	**40** 1997- 2002

Author	Venue	Cited	#Paper
6 Jiawei Han Xifeng Yan	**1** SIG- MOD	**1** p76095[2]	**22** 2004- 2010

[2] "Frequent subgraph discovery"

Venue	Keyword	#Paper
3 ICDM AAAI TKDE	**1** "anomaly detection"	**25** 2005- 2013

Author	Venue	Keyword	#Paper
27 C. Faloutsos J. Pei P. S. Yu X. Lin C. Aggarwal...	**6** KDD ICDM ICDE TKDE …	**12** "large graphs" "data streams" "evolving data" "evolving graphs" …	**70** 2006- 2013

Author	Venue	Keyword	Cited	#Paper
12 Ryen White Hang Li Tie-Yan Liu Zhaohui Zheng…	**5** SIGIR WWW WSDM CIKM…	**3** "web search" "click-through data" "sponsored search"	**12** p82630[3] p116290 p103899 p106191…	**32** 2006- 2013

[3] "Optimizing search engines using clickthrough data"

Author	Venue	Keyword	#Paper
8 Qiang Yang Dou Shen Sinno Pan...	**3** KDD PAKDD AAAI	**6** "transfer learning" "data mining" "localization models"	**17** 2007- 2010

Jiang, et al. (KDD 2016)

Summary

- Density-based methods and applications
  - Density in temporal bipartite graphs
    - CopyCatch (WWW'13): Facebook, ill-gotten likes
  - Density (synchronicity) in large directed graphs
    - CatchSync (KDD'14): Twitter/Weibo, zombie followers
  - Density (suspiciousness) in multidimensional data
    - CrossSpot (TKDE'16): Twitter/Weigo, social spam
  - Density (MDL principle) in multicontextual data
    - CatchTartan (KDD'16): Twitter, events/campaigns

References

- Meng Jiang, Peng Cui, and Christos Faloutsos. “Suspicious behavior detection: current trends and future directions.” IEEE Intelligent Systems, 2016. (Survey paper)
- Alex Beutel, Wanhong Xu, Venkatesan Guruswami, Christopher Palow, Christos Faloutsos. “Copycatch: stopping group attacks by spotting lockstep behavior in social networks”, WWW 2013.
- Meng Jiang, Peng Cui, Alex Beutel, Christos Faloutsos, Shiqiang Yang. “CatchSync: Catching Synchronized Behavior in Large Directed Graphs”, KDD 2014 Best Paper Finalist.
- Meng Jiang, Alex Beutel, Peng Cui, Bryan Hooi, Shiqiang Yang, Christos Faloutsos. “Spotting Suspicious Behaviors in Multimodal Data: A General Metric and Algorithms”, TKDE 2016.
- Meng Jiang, Christos Faloutsos, Jiawei Han. “CATCHTARTAN: Representing and Summarizing Dynamic Multicontextual Behaviors”, KDD 2016.
- M. Faloutsos, P. Faloutsos, and C. Faloutsos. “On power-law relationships of the internet topology.” SIGCOMM, 1999.
- A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Weiner. “Graph structure in the web.” Computer Networks, 2000.
- F. Chung and L. Lu. “The average distances in random graphs with given expected degrees.” PNAS, 2002.
- J. Kleinberg. “Authoritative sources in a hyperlinked environment.” JACM, 1999.
- H. Kwak, C. Lee, H. Park, and S. Moon. “What is Twitter, a social network or a news media?” WWW, 2010.

KDD 2017
Halifax, Nova Scotia - Canada
August 13 - 17, 2017

Tutorial: Data-Driven Approaches towards Malicious Behavior Modeling

Meng Jiang
University of Notre Dame

Srijan Kumar
Stanford University

Christos Faloutsos
Carnegie Mellon University

V.S. Subrahmanian
University of Maryland, College Park

Tutorial link: http://bit.ly/kdd2017