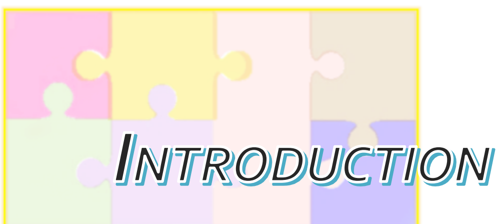
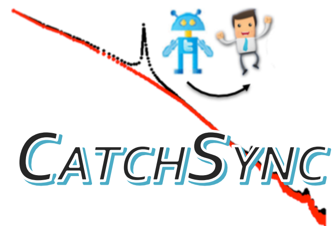
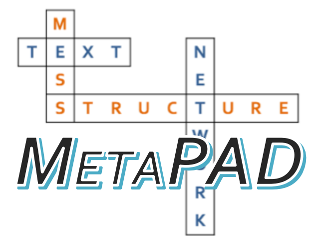
Data-Driven Behavioral Analytics with Networks







Meng Jiang
University of Notre Dame

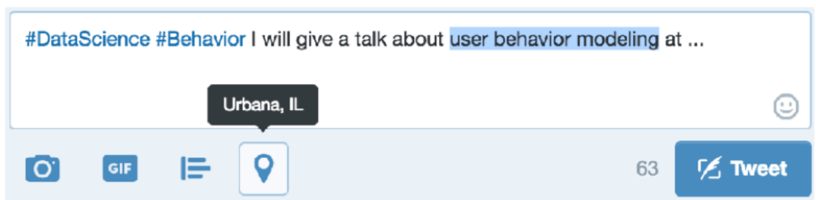
http://www.meng-jiang.com

Behavior and "Behavior Networks"

Interaction of individuals with themselves and with their environment. —Wikipedia

Social behaviors

Tweeting behaviors



Paper-publishing behaviors

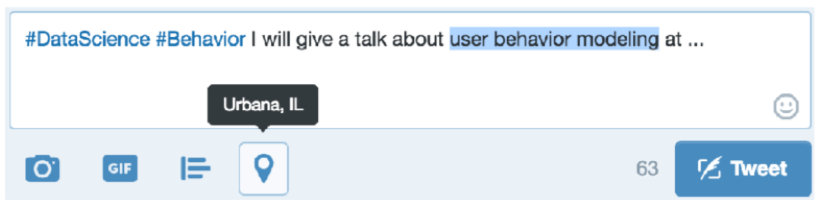
Meng Jiang, **Christos Faloutsos**, and **Jiawei Han**. "CatchTartan: **Representing** and **Summarizing Dynamic Behaviors**." In **SIGKDD 2016**.

Behavior and "Behavior Networks"

Interaction of individuals with themselves and with their environment. —Wikipedia

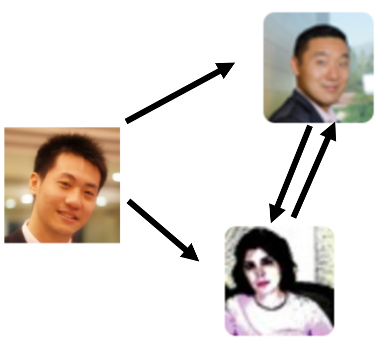
Social behaviors

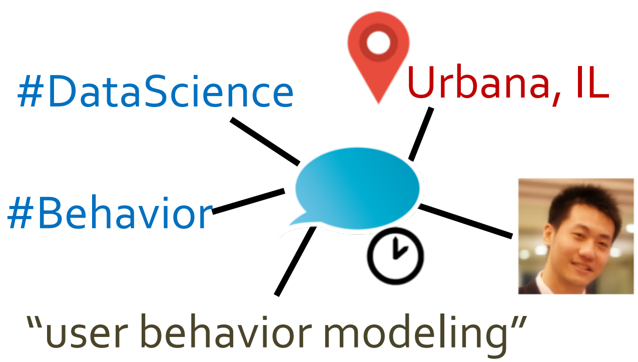
Tweeting behaviors



Paper-publishing behaviors

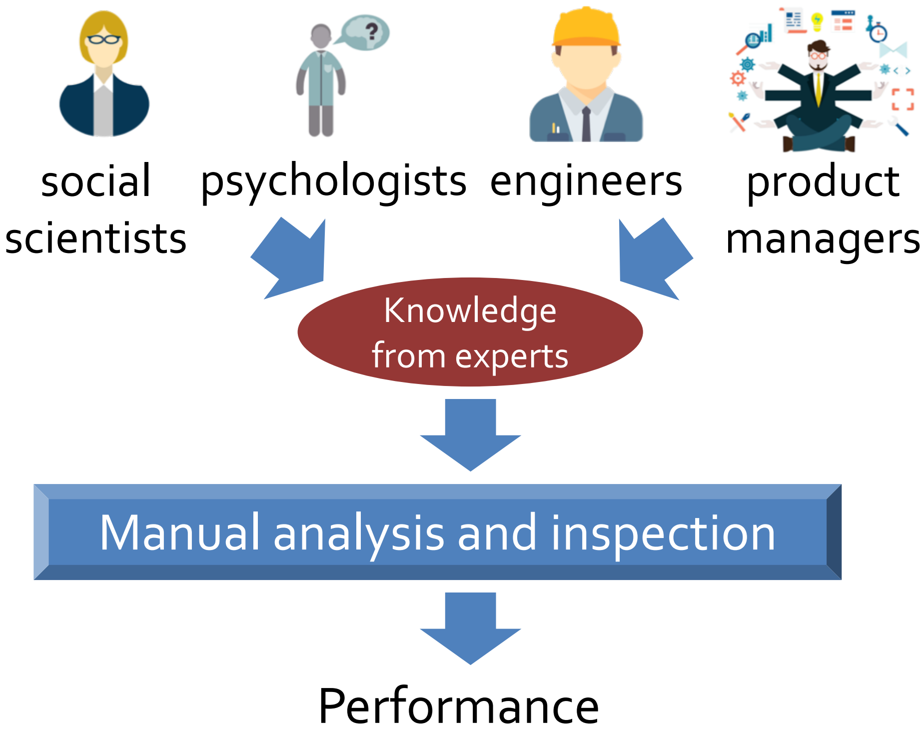
Meng Jiang, **Christos Faloutsos**, and **Jiawei Han**. "CatchTartan: **Representing** and **Summarizing Dynamic Behaviors**." In **SIGKDD 2016**.



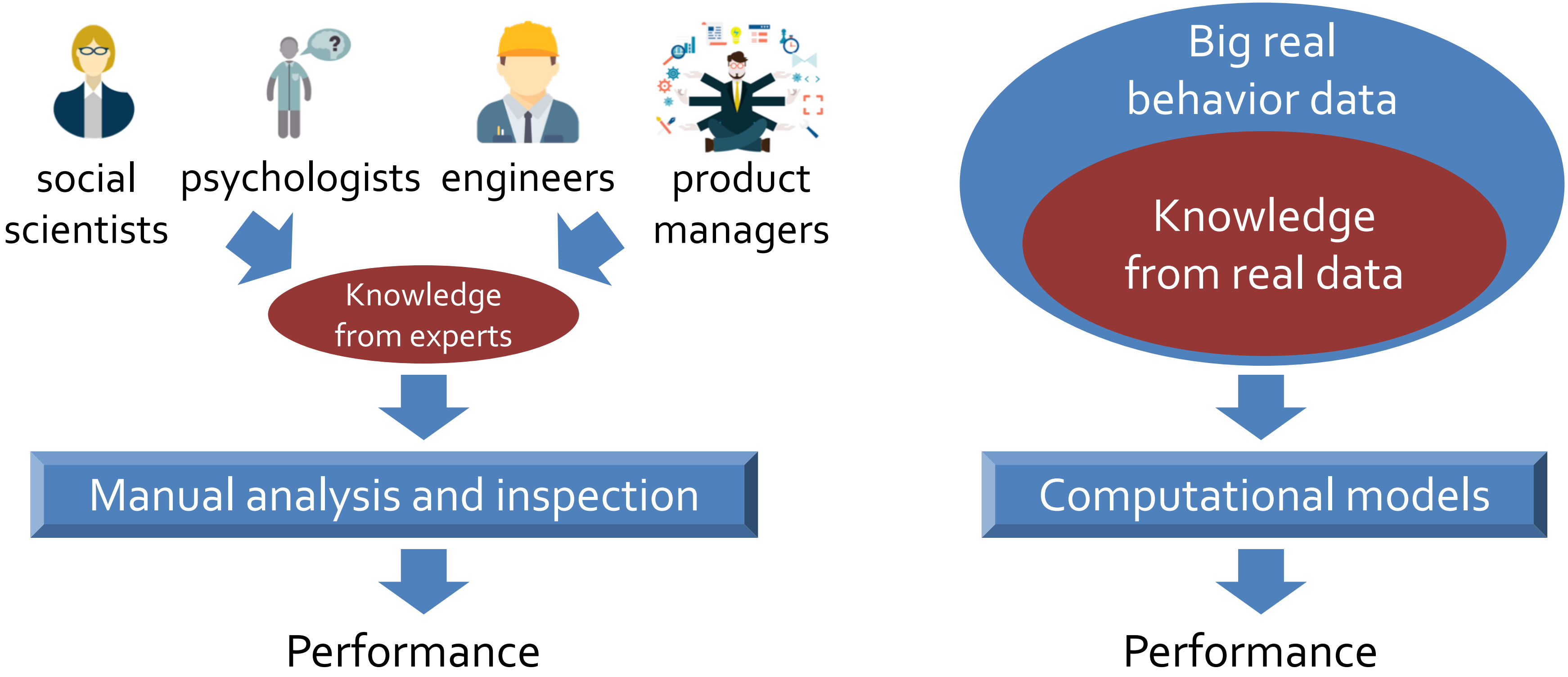


Applications: prediction, recommendation, fraud detection, spam detection…

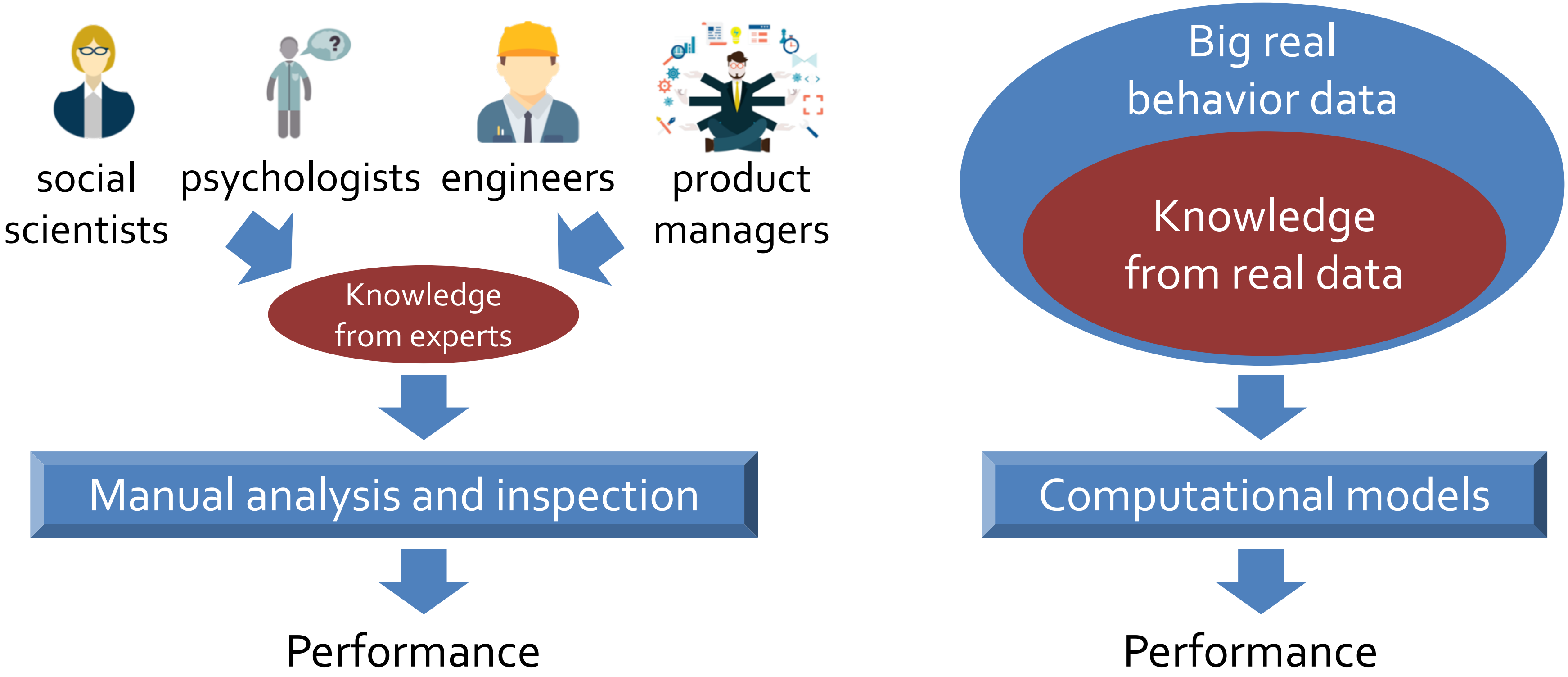
Data-Driven Behavioral Analysis



Data-Driven Behavioral Analysis



Data-Driven Behavioral Analysis



Data, knowledge, intelligence, and trustworthiness.

(User behavior modeling)

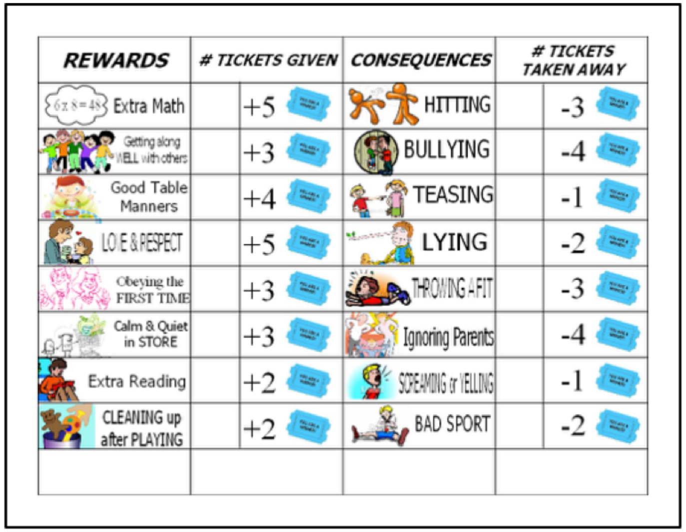
Basic Research Areas

- Six Disruptive Basic Research Areas
  - Engineered Materials (metamaterials and plasmonics)
  - Quantum Information and Control
  - Cognitive Neuroscience
  - Nanoscience and Nanoengineering
  - Synthetic Biology
  - Computational Modeling of Human and Social Behavior

Research Topics in Behavior Modeling

Behavior Modeling

Research Topics in Behavior Modeling

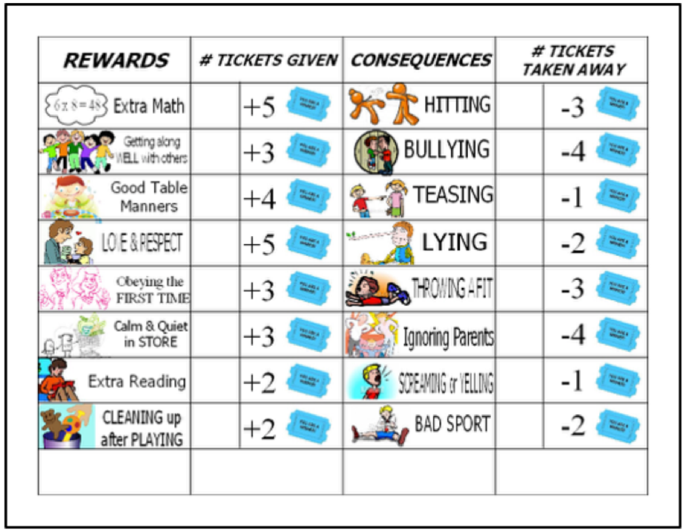


REWARDS	# TICKETS GIVEN	CONSEQUENCES	# TICKETS TAKEN AWAY
6 x 8 = 48 Extra Math	+5	HITTING	-3
Getting along WELL with others	+3	BULLYING	-4
Good Table Manners	+4	TEASING	-1
LOVE & RESPECT	+5	LYING	-2
Obeying the FIRST TIME	+3	THROWING A FIT	-3
Calm & Quiet in STORE	+3	Ignoring Parents	-4
Extra Reading	+2	SCREAMING or YELLING	-1
CLEANING up after PLAYING	+2	BAD SPORT	-2

1. Behavior intentions

Behavior Modeling

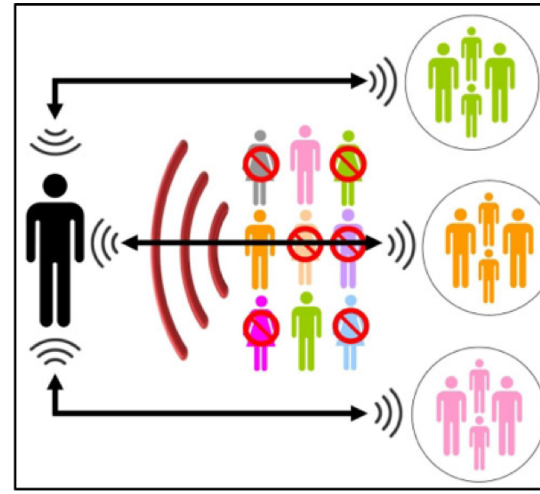
Research Topics in Behavior Modeling



REWARDS	# TICKETS GIVEN	CONSEQUENCES	# TICKETS TAKEN AWAY
Extra Math	+5	HITTING	-3
Getting along WELL with others	+3	BULLYING	-4
Good Table Manners	+4	TEASING	-1
LOVE & RESPECT	+5	LYING	-2
Obeying the FIRST TIME	+3	THROWING A FIT	-3
Calm & Quiet in STORE	+3	Ignoring Parents	-4
Extra Reading	+2	SCREAMING or YELLING	-1
CLEANING up after PLAYING	+2	BAD SPORT	-2

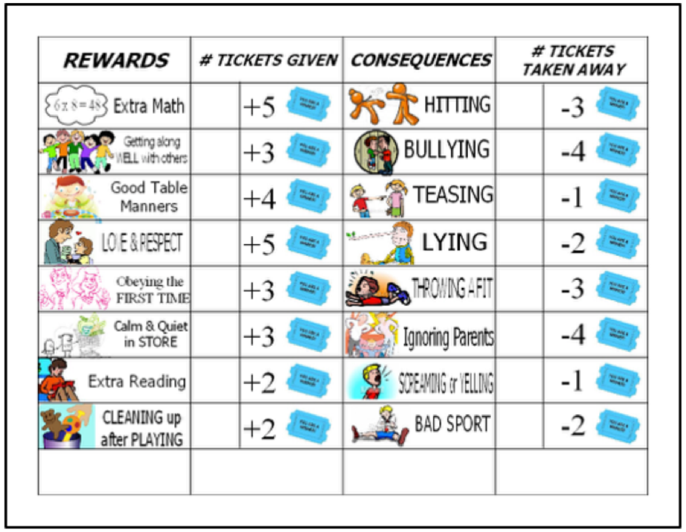
1. Behavior intentions

2. Social contexts



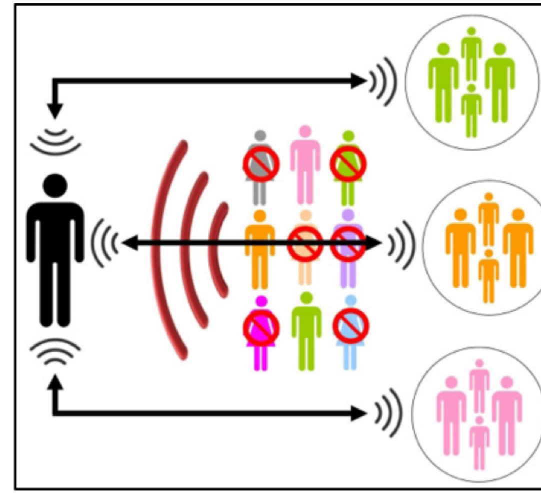
Behavior Modeling

Research Topics in Behavior Modeling



REWARDS	# TICKETS GIVEN	CONSEQUENCES	# TICKETS TAKEN AWAY
Extra Math	+5	HITTING	-3
Getting along WELL with others	+3	BULLYING	-4
Good Table Manners	+4	TEASING	-1
LOVE & RESPECT	+5	LYING	-2
Obeying the FIRST TIME	+3	THROWING A FIT	-3
Calm & Quiet in STORE	+3	Ignoring Parents	-4
Extra Reading	+2	SCREAMING or YELLING	-1
CLEANING up after PLAYING	+2	BAD SPORT	-2

1. Behavior intentions

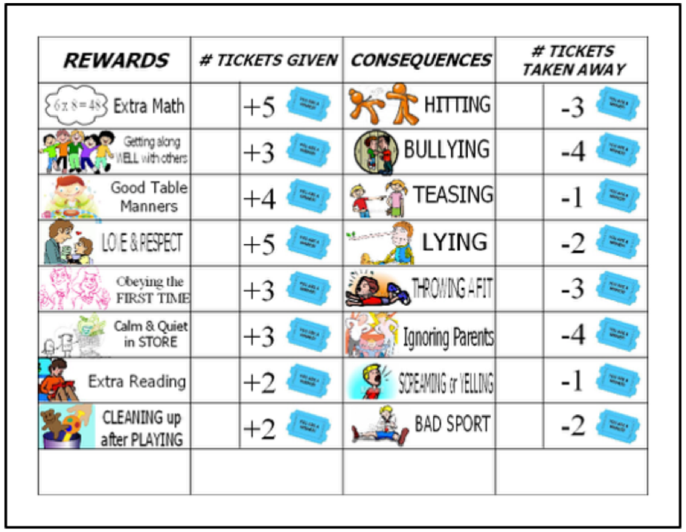


2. Social contexts

Behavior Modeling

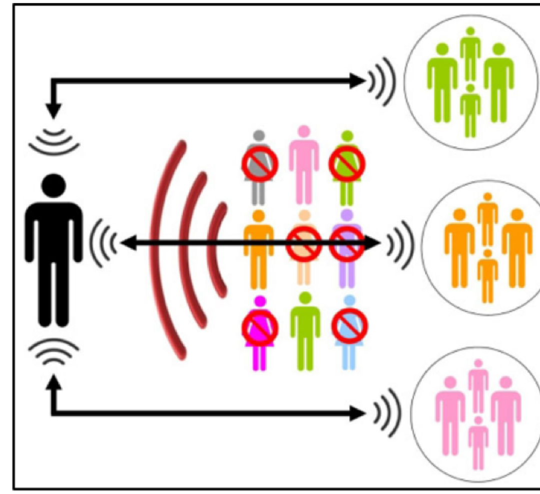
3. Spatiotemporal contexts

Research Topics in Behavior Modeling



REWARDS	# TICKETS GIVEN	CONSEQUENCES	# TICKETS TAKEN AWAY
Extra Math	+5	HITTING	-3
Getting along WELL with others	+3	BULLYING	-4
Good Table Manners	+4	TEASING	-1
LOVE & RESPECT	+5	LYING	-2
Obeying the FIRST TIME	+3	THROWING A FIT	-3
Calm & Quiet in STORE	+3	Ignoring Parents	-4
Extra Reading	+2	SCREAMING or YELLING	-1
CLEANING up after PLAYING	+2	BAD SPORT	-2

1. Behavior intentions

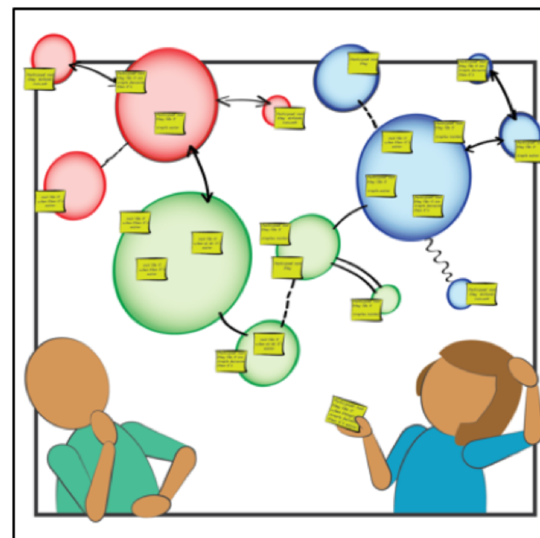


2. Social contexts

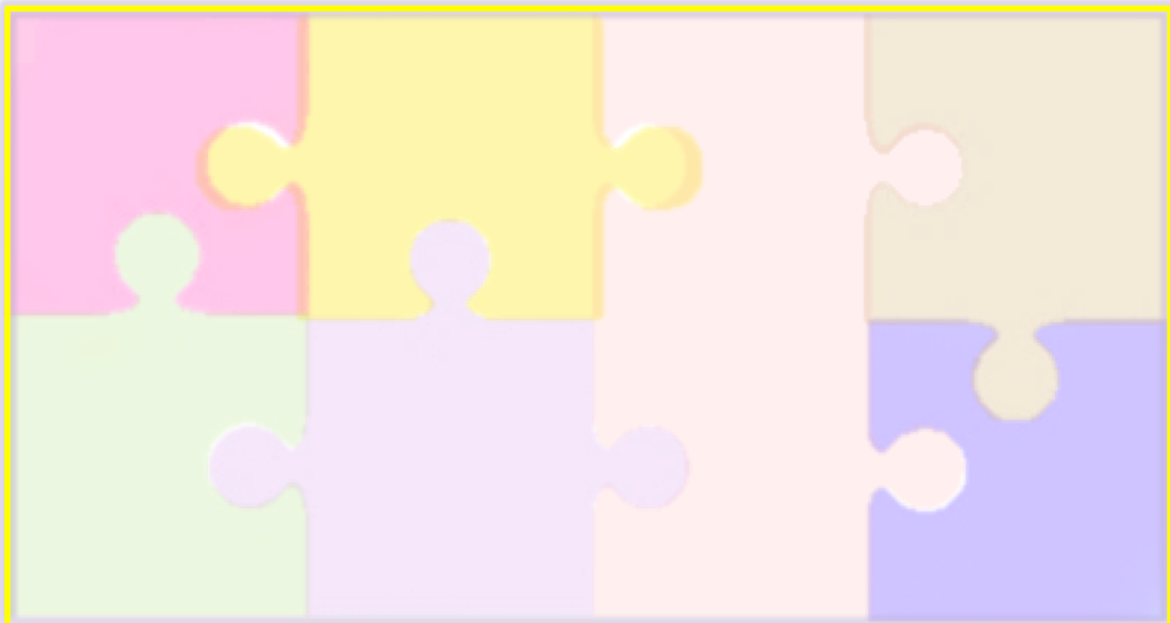
Behavior Modeling

3. Spatiotemporal contexts

4. Behavior content



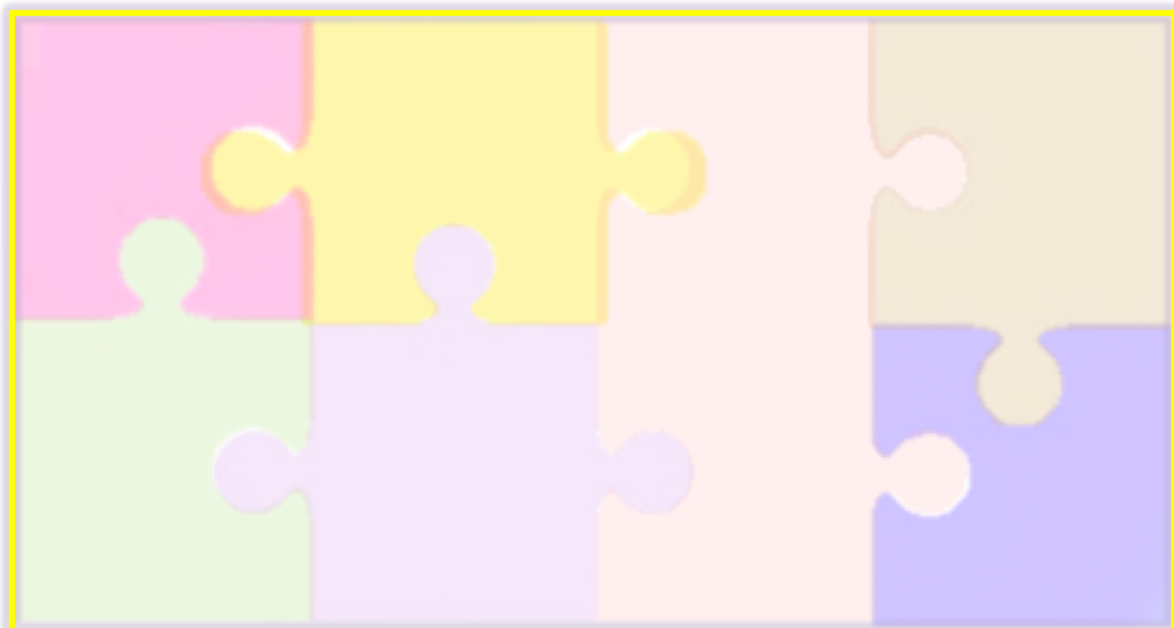
My Research "Area"



My Research "Area"

Intelligence:
Behavior prediction and recommendation

Trustworthiness:
Suspicious behavior detection



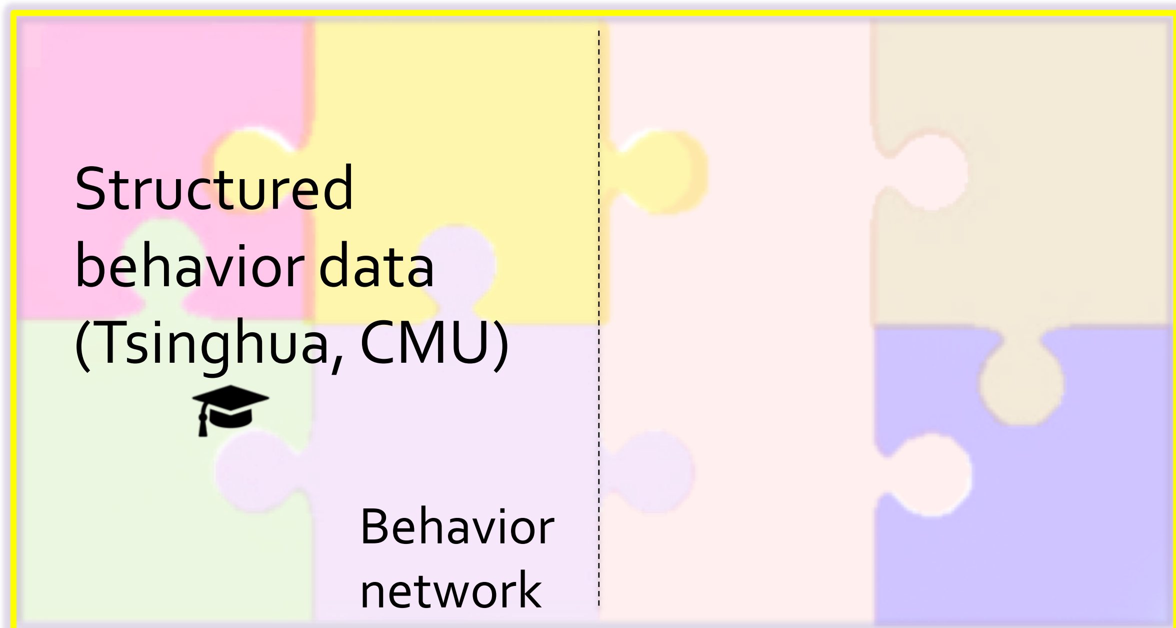
My Research “Area”

Social contexts

Spatiotemporal contexts

Intelligence:
Behavior prediction and recommendation

Trustworthiness:
Suspicious behavior detection



My Research “Area”

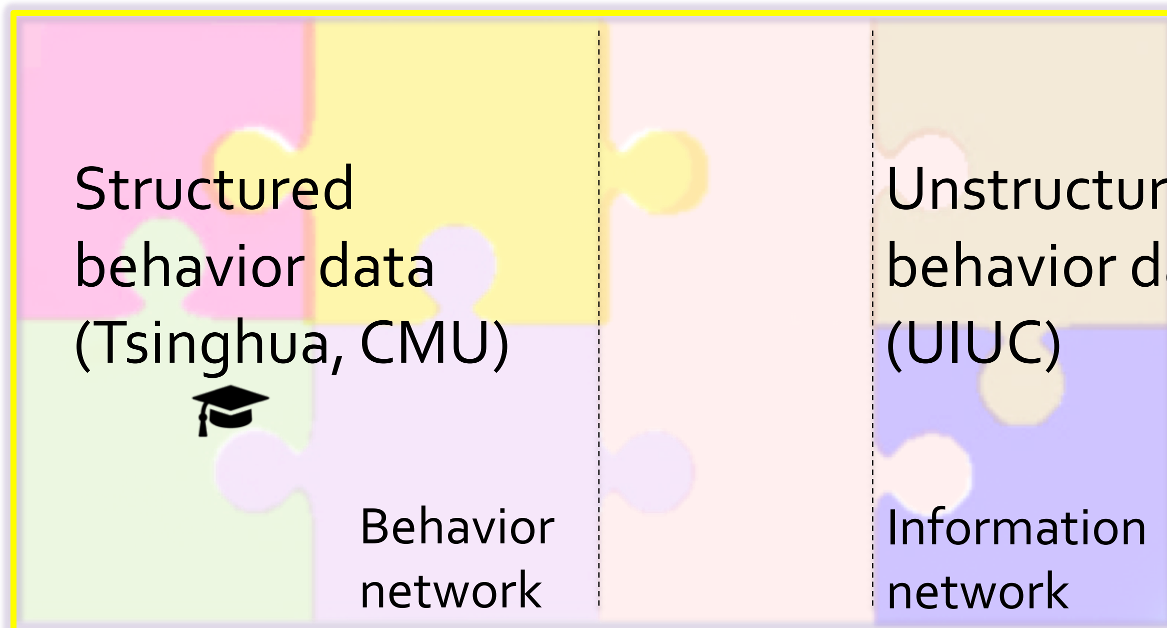
Social contexts

Spatiotemporal contexts

Behavioral content

Intelligence:
Behavior prediction and recommendation

Trustworthiness:
Suspicious behavior detection

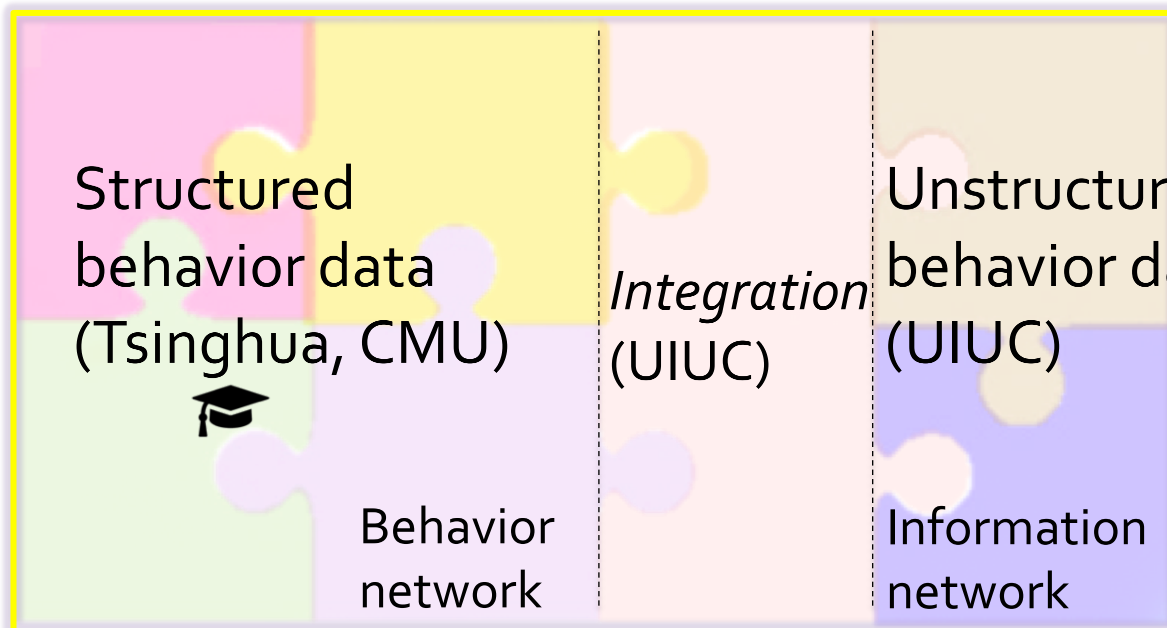


My Research “Area”

Intelligence:
Behavior prediction and recommendation

Trustworthiness:
Suspicious behavior detection

Social contexts | Spatiotemporal contexts | Behavioral content



My Research "Area"

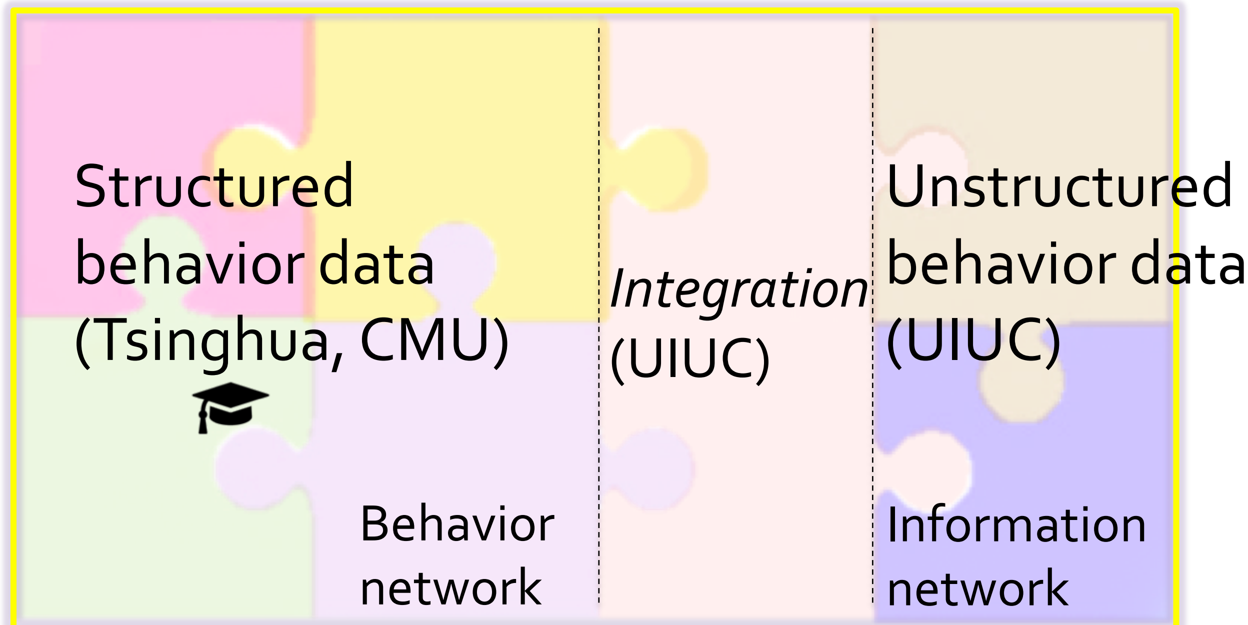
Intelligence:
Behavior prediction and recommendation

Trustworthiness:
Suspicious behavior detection

Social contexts

Spatiotemporal contexts

Behavioral content



Ask good **Q**uestions.

Find good Data-Driven **M**ethodologies.

Propose good **S**olutions.

Q1: A Social Recommender System

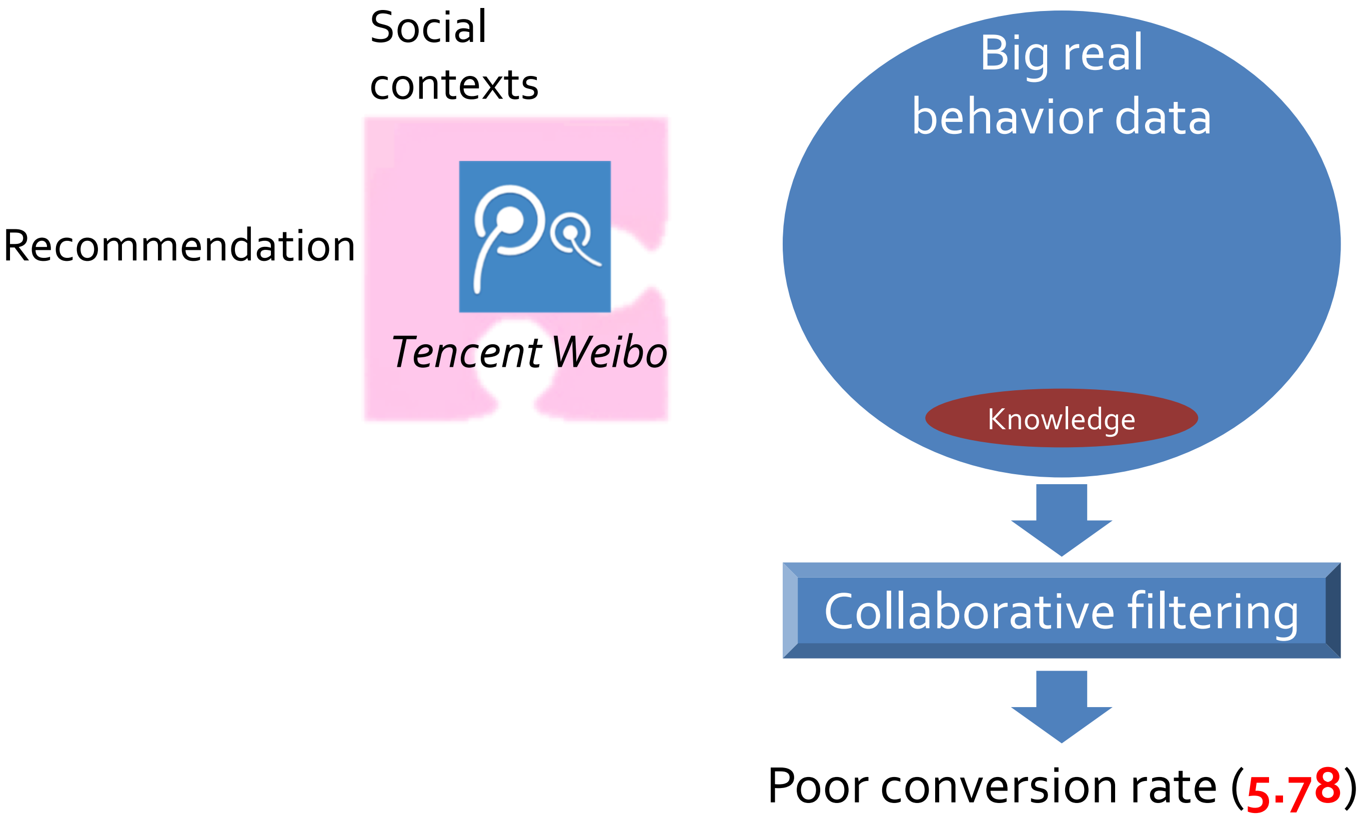
Social contexts

Recommendation

Tencent 腾讯

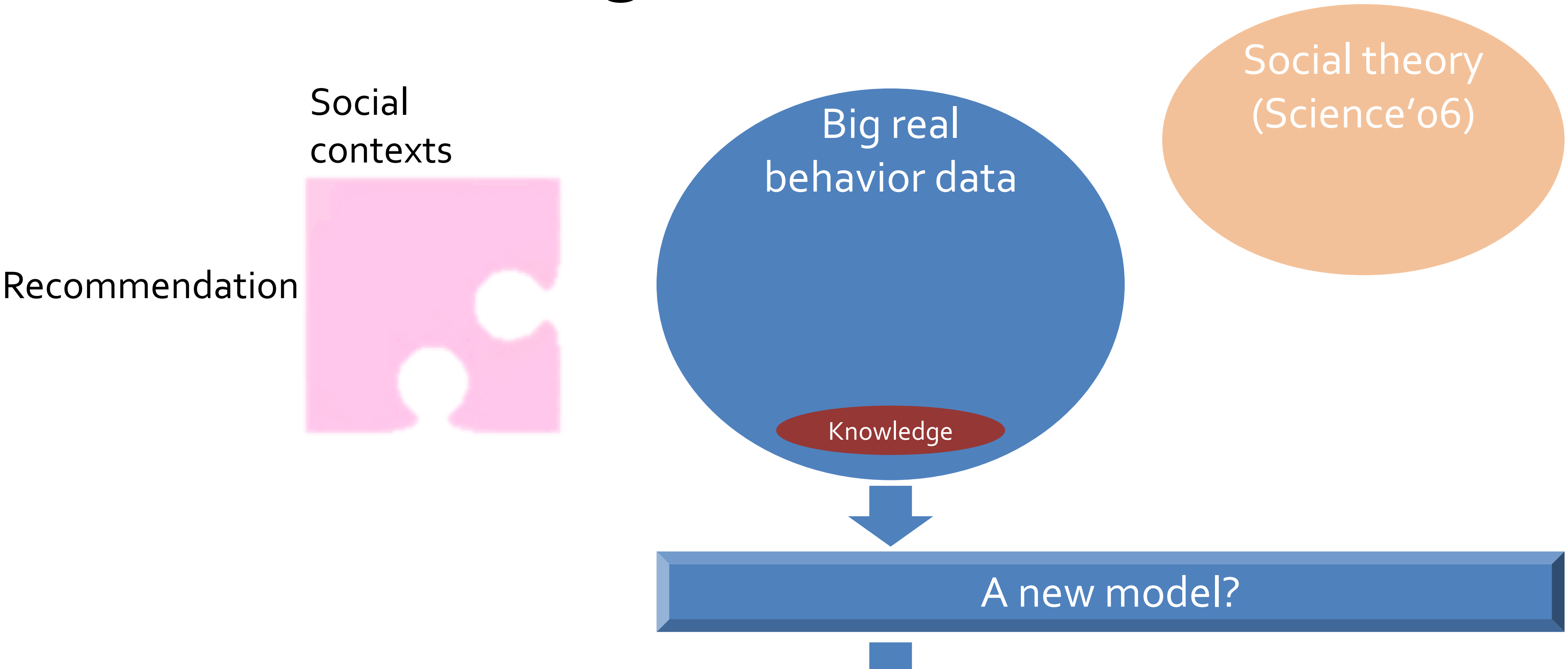
清華大學
Tsinghua University

Q1: A Social Recommender System



Collaborative filtering for recommenders [Breese et al. UAI'98; Getoor and Sahami WEBKDD'99; Herlocker et al. CSCW'00, TOIS'04; Koren et al. KDD'08 Computer'09; Liu et al. SIGIR'08]

M1: Knowledge from Social Theories



"… Differences between **<u>independent</u>** and **<u>social influence conditions</u>** are significant …"

Salganik, Dodds, and Watts. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market." Science, Vol. 311, 2006.

M1: Knowledge from Social Theories

Social contexts

Recommendation

Big real behavior data

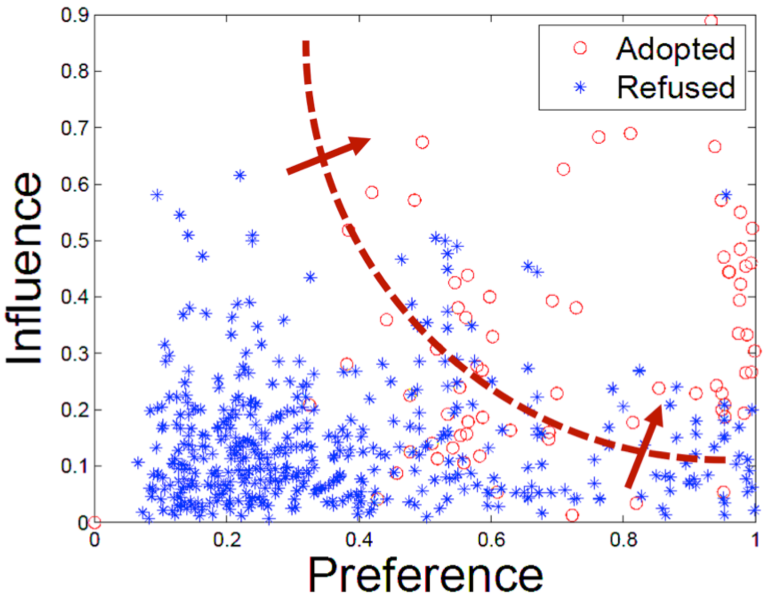
Knowledge

Social theory (Science'06)

Observation from data

A new model?

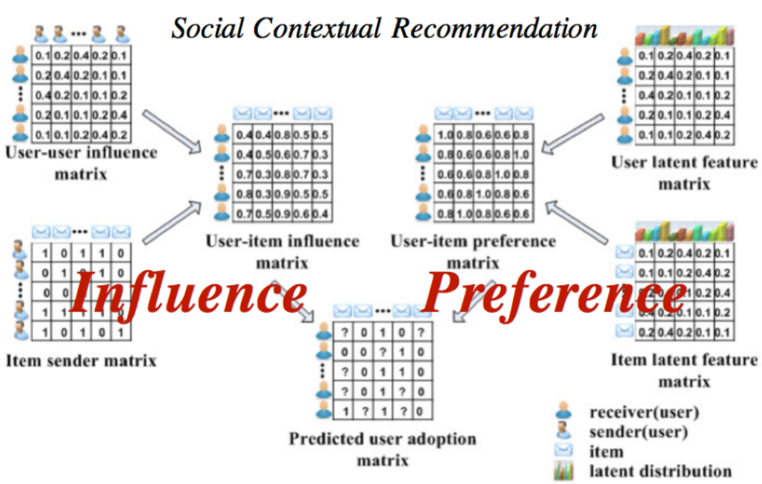
Performance?

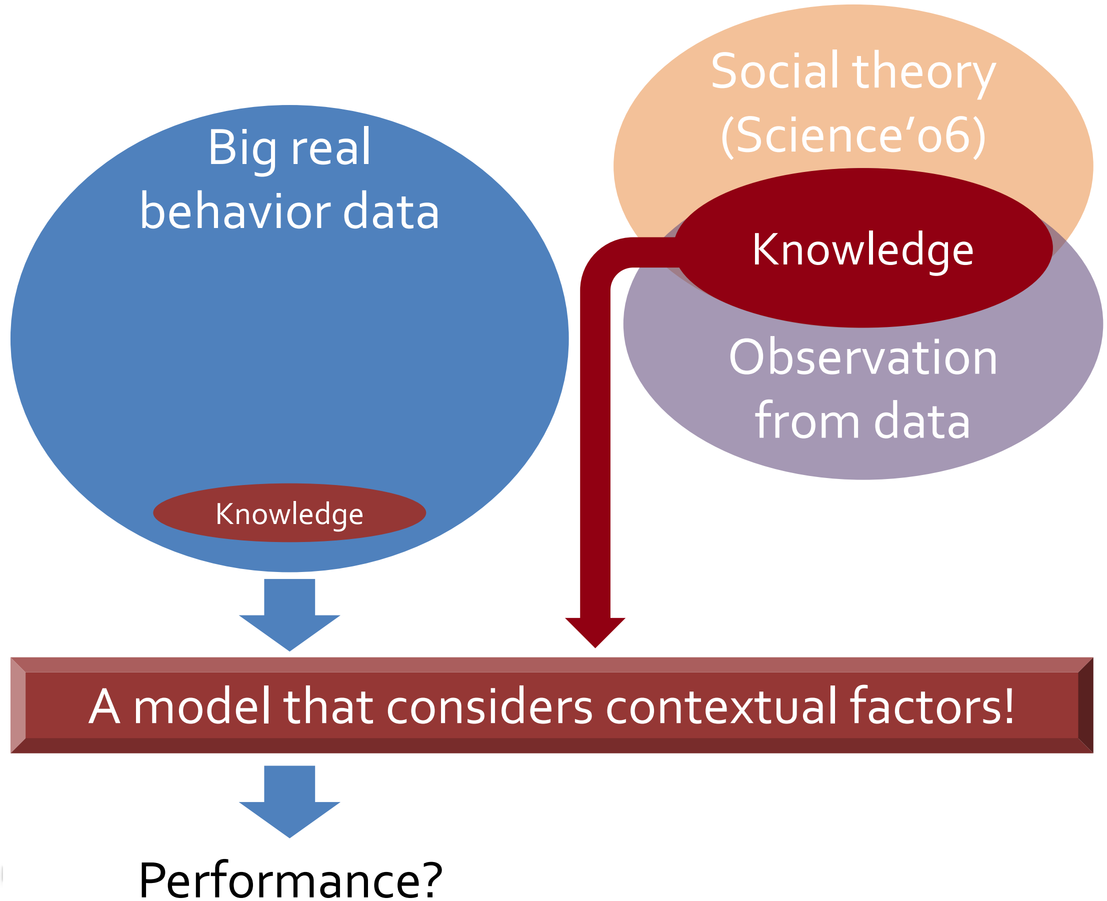


M1: Knowledge from Social Theories

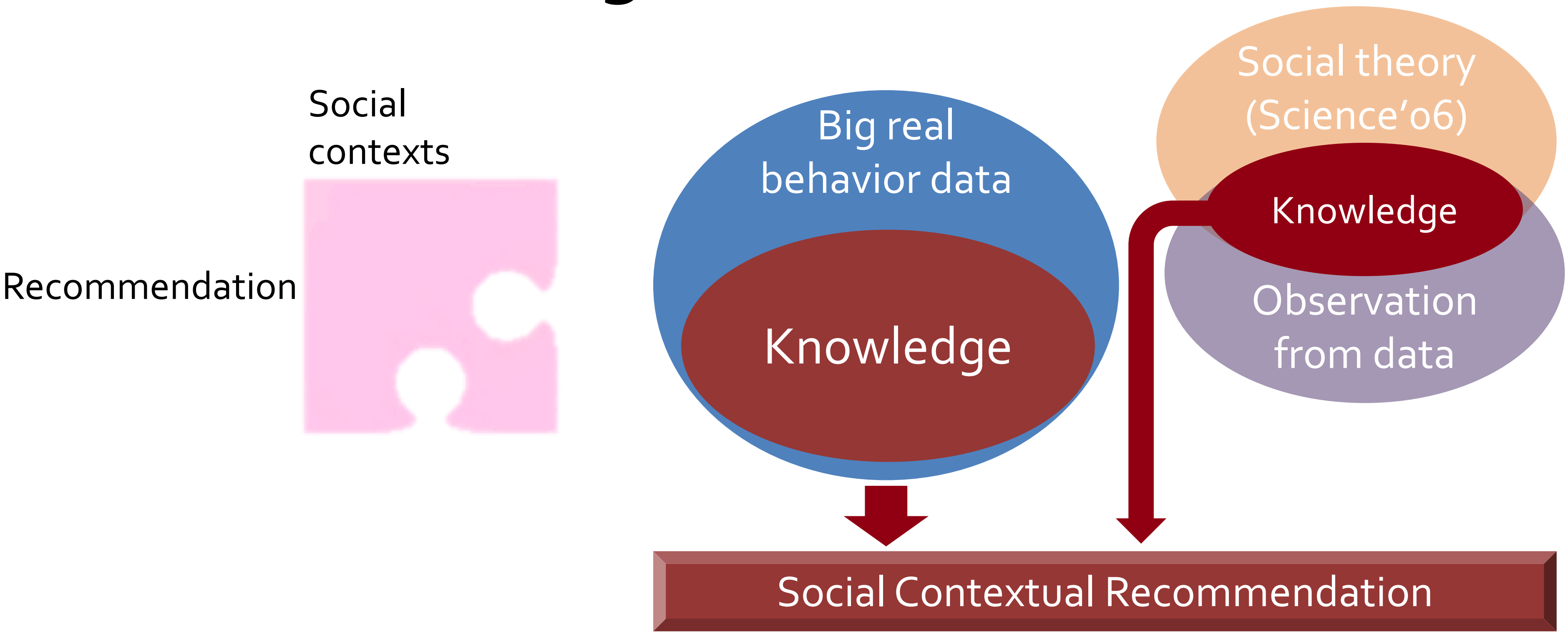
Social contexts

Recommendation





M1: Knowledge from Social Theories



Higher conversion rate (**8.27, relatively +43%**)

S1: CIKM'12: cited by **211**; TKDE'14: cited by **82**.

S1: Social Contextual Recommendation

- Optimization

behavior influence preference

$$P(\mathbf{R}|\mathbf{S},\mathbf{U},\mathbf{V},\sigma_R^2) = \prod_{i=1}^{M}\prod_{j=1}^{N}\mathcal{N}(\underline{\mathbf{R}_{ij}}|\underline{\mathbf{S}_i\mathbf{G}_j^\top}\odot\underline{\mathbf{U}_i^\top\mathbf{V}_j},\sigma_R^2)$$

behavior user-user interaction

$$\mathcal{J} = ||\mathbf{R}-\mathbf{S}\mathbf{G}^\top\odot\mathbf{U}^\top\mathbf{V}||_F^2 + \alpha||\mathbf{W}-\mathbf{U}^\top\mathbf{U}||_F^2$$
$$+\beta||\mathbf{C}-\mathbf{V}^\top\mathbf{V}||_F^2 + \gamma||\mathbf{S}-\mathbf{F}||_F^2$$
$$+\delta||\mathbf{S}||_F^2 + \eta||\mathbf{U}||_F^2 + \lambda||\mathbf{V}||_F^2$$

social relation

item content

avoid overfitting

- Gradient descent method

$$\frac{\partial\mathcal{J}}{\partial\mathbf{S}} = 2\Big(-\mathbf{R}(\mathbf{G}\odot\mathbf{V}^\top\mathbf{U}) + (\mathbf{S}\mathbf{G}^\top\odot\mathbf{U}^\top\mathbf{V})\mathbf{G} + \gamma(\mathbf{S}-\mathbf{F}) + \delta\mathbf{S}\Big)$$

$$\frac{\partial\mathcal{J}}{\partial\mathbf{U}} = 2\Big(-\mathbf{V}\mathbf{R}^\top + \mathbf{V}(\mathbf{G}\mathbf{S}^\top\odot\mathbf{V}^\top\mathbf{U}) - 2\alpha\mathbf{U}\mathbf{W} + 2\alpha\mathbf{U}\mathbf{U}^\top\mathbf{U} + \eta\mathbf{U}\Big)$$

$$\frac{\partial\mathcal{J}}{\partial\mathbf{V}} = 2\Big(-\mathbf{U}\mathbf{R} + \mathbf{U}(\mathbf{S}\mathbf{G}^\top\odot\mathbf{U}^\top\mathbf{V}) - 2\beta\mathbf{V}\mathbf{C} + 2\beta\mathbf{V}\mathbf{V}^\top\mathbf{V} + \lambda\mathbf{V}\Big)$$

Q2: Leveraging Social Data for Prediction

Social contexts

UBER NETFLIX airbnb

Behavior prediction

Sparse data on target platforms

Knowledge

Predictive models

Poor accuracy

Data sparseness [Herlocker et al. CSCW'00; Sarwar et al. WWW'01; Burke UM&UAI'02; Ma et al. TOIS'11 TIST'11; Tang et al. Soc. Netw. Anal. Min.; Xue et al. VLDB'15; Han and Obradovic et al. SDM'16]

Q2: Leveraging Social Data for Prediction

Social contexts

Behavior prediction

UBER airbnb NETFLIX

Sparse data on target platforms

Knowledge

Predictive models

Poor accuracy

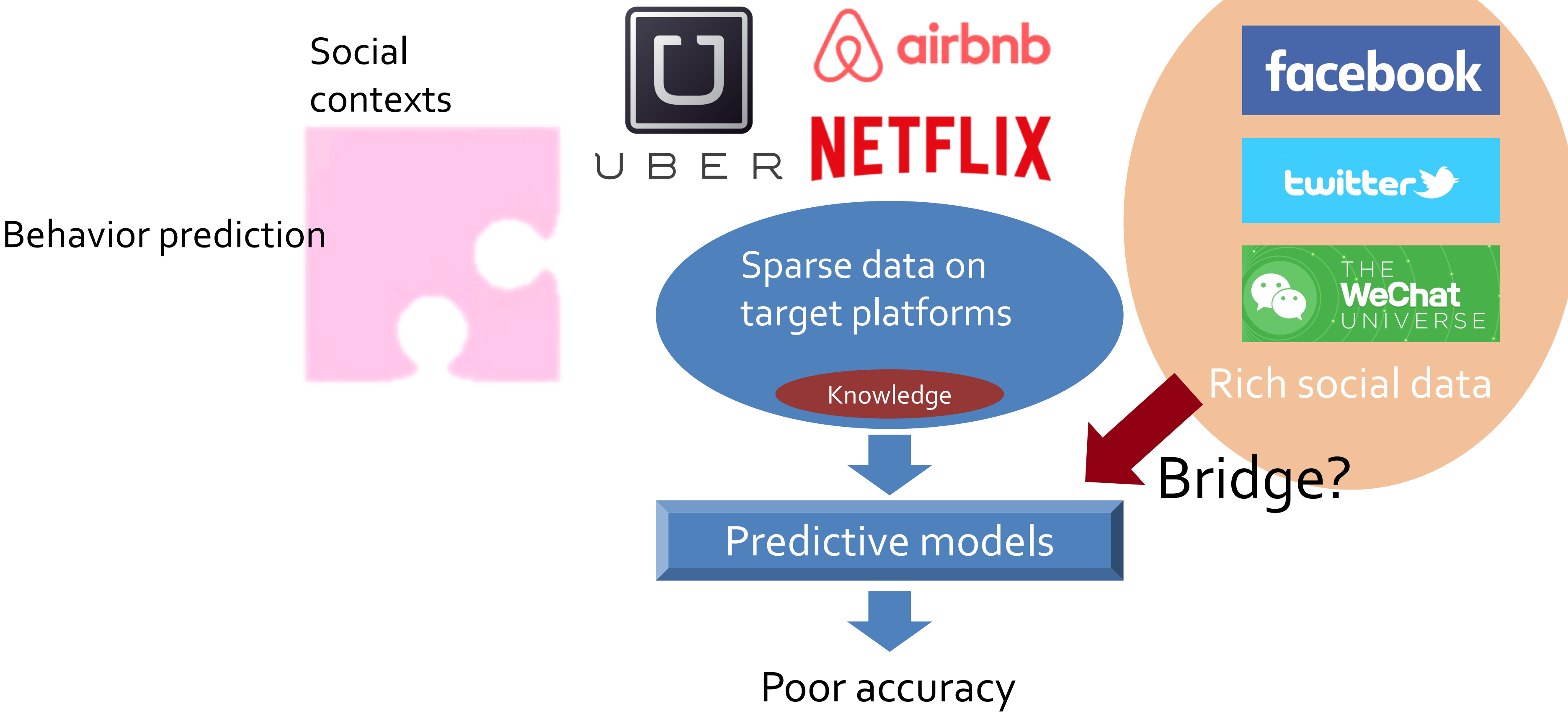
facebook

twitter

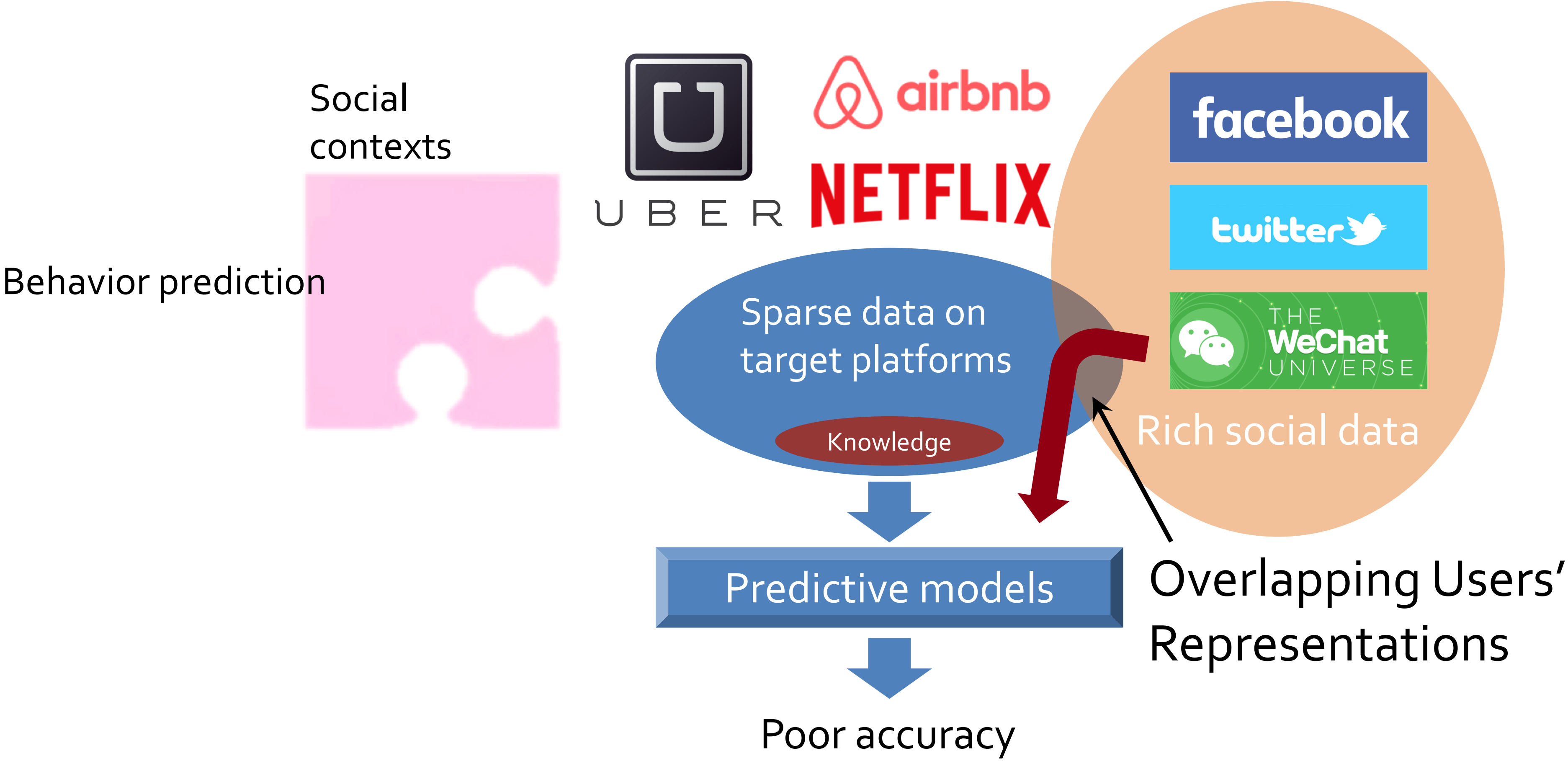
THE WeChat UNIVERSE

Rich social data

M2: Knowledge Transfer from Heterogeneous Sources



M2: Knowledge Transfer from Heterogeneous Sources



M2: Knowledge Transfer from Heterogeneous Sources

Social contexts

Behavior prediction

~50K users

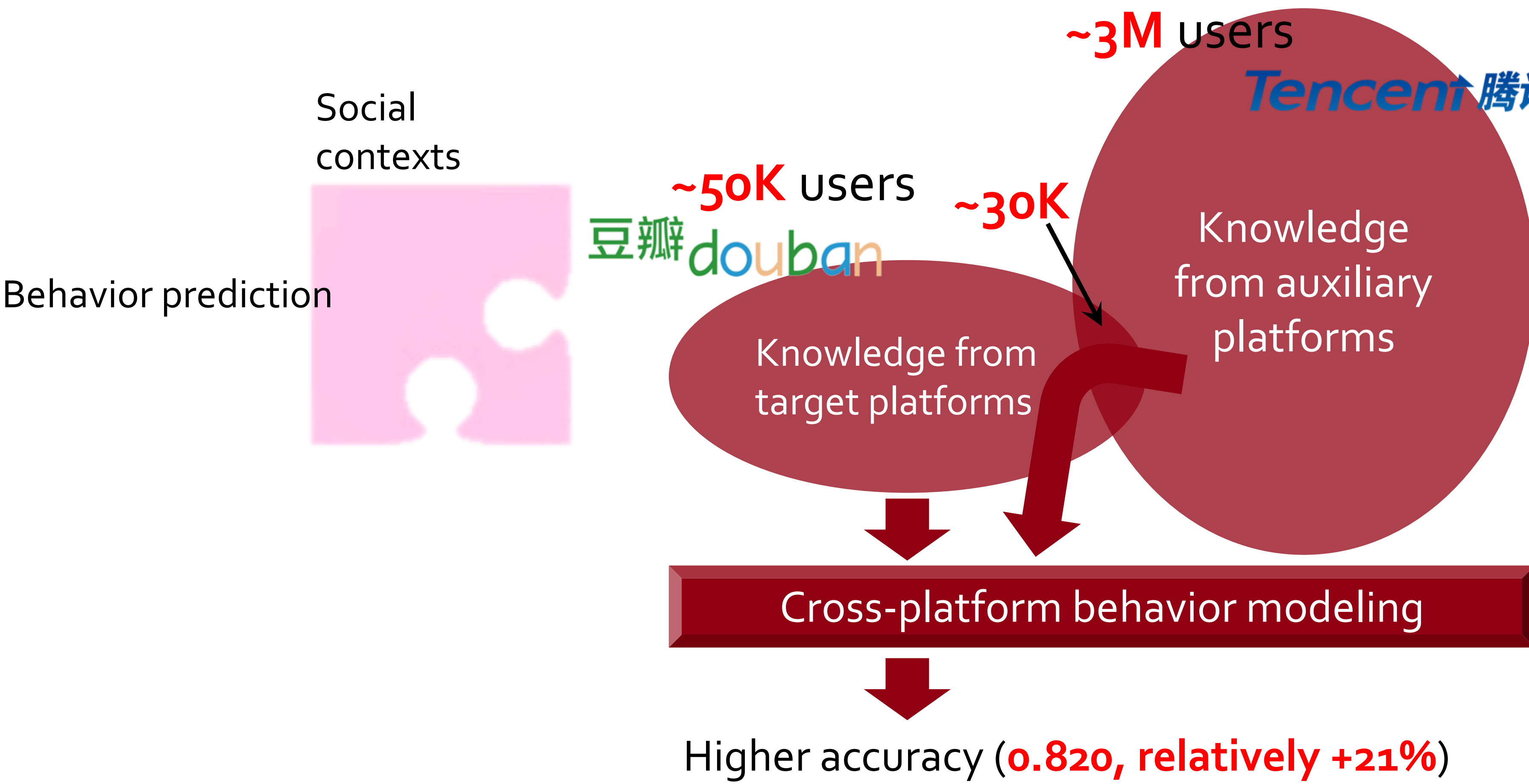
豆瓣 douban

Knowledge from target platforms

Predictive models

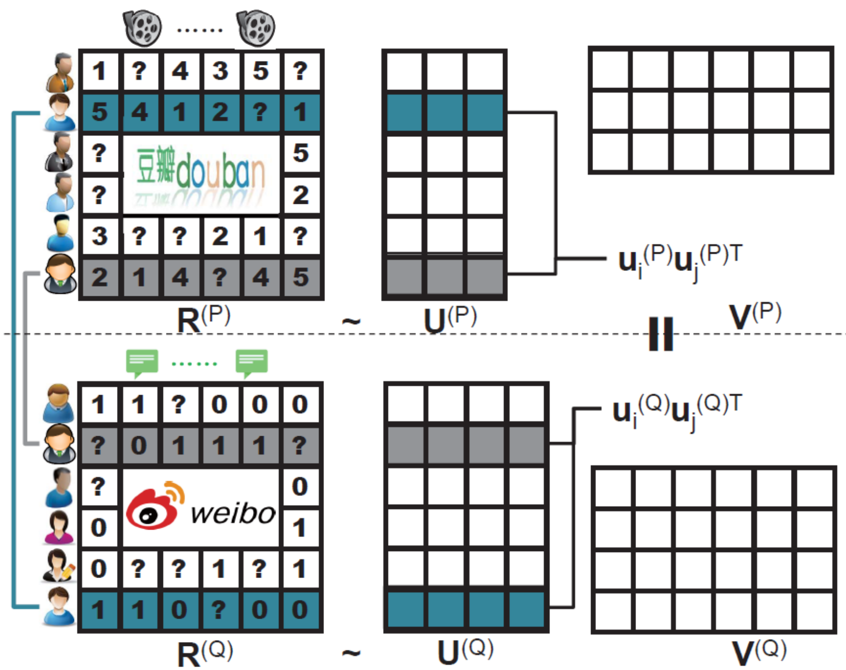
Poor accuracy (0.676)

M2: Knowledge Transfer from Heterogeneous Sources



S2: CIKM’12: cited by **72**; TKDE’15: cited by **44**; AAAI’16: cited by **13**.

S2: Cross-Platform Behavior Modeling



Target platform

Auxiliary platform

$$
\begin{aligned}
\mathcal{J} = & \sum_{i,j} W_{i,j}^{(P)} \left(R_{i,j}^{(P)} - \sum_{r} U_{i,r}^{(P)} V_{r,j}^{(P)} \right)^2 \\
& + \lambda \sum_{i,j} W_{i,j}^{(Q)} \left(R_{i,j}^{(Q)} - \sum_{r} U_{i,r}^{(Q)} V_{r,j}^{(Q)} \right)^2 \\
& + \mu \sum_{i_1,j_1,i_2,j_2} W_{i_1,j_1}^{(P,Q)} W_{i_2,j_2}^{(P,Q)} \left(A_{i_1,i_2}^{(P)} - A_{j_1,j_2}^{(Q)} \right)^2
\end{aligned}
$$

Overlapping user similarity
(Pair-wise regularization)

Q3: Catching Social Link Farming

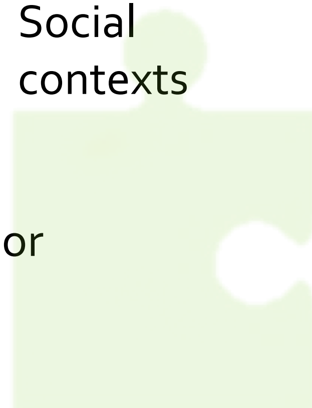
Social contexts

Suspicious behavior detection

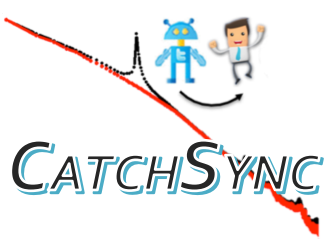
Meng Jiang, Peng Cui, and Christos Faloutsos. "Suspicious behavior detection: current trends and future directions." **IEEE Intelligent Systems**, 2016. (Survey paper)

Q3: Catching Social Link Farming

Social contexts



Suspicious behavior detection

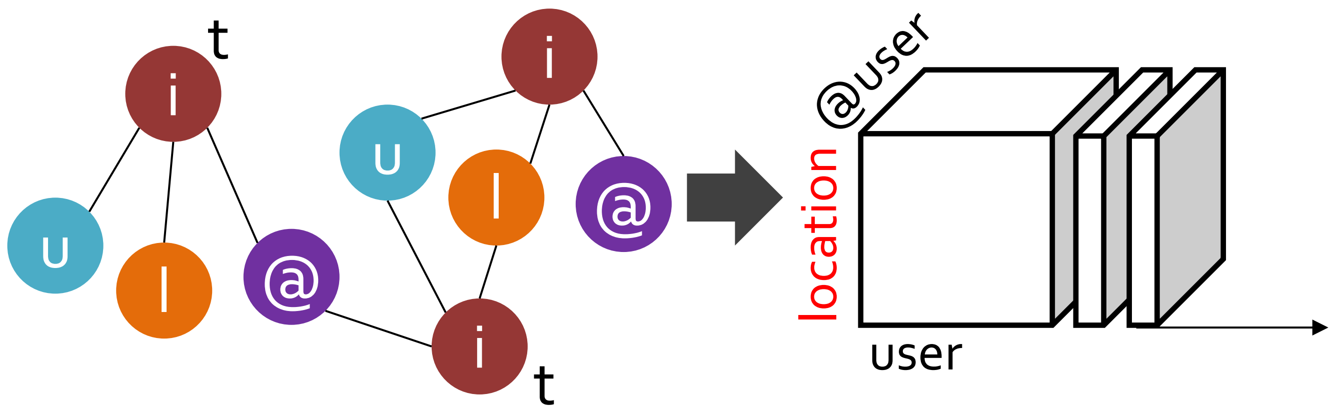
S3: 

KDD'14 best paper finalist

Q4: Knowledge from Spatiotemporal Information
M4: Tensor Methods for Modeling Multiple Dimensions

Spatiotemporal contexts

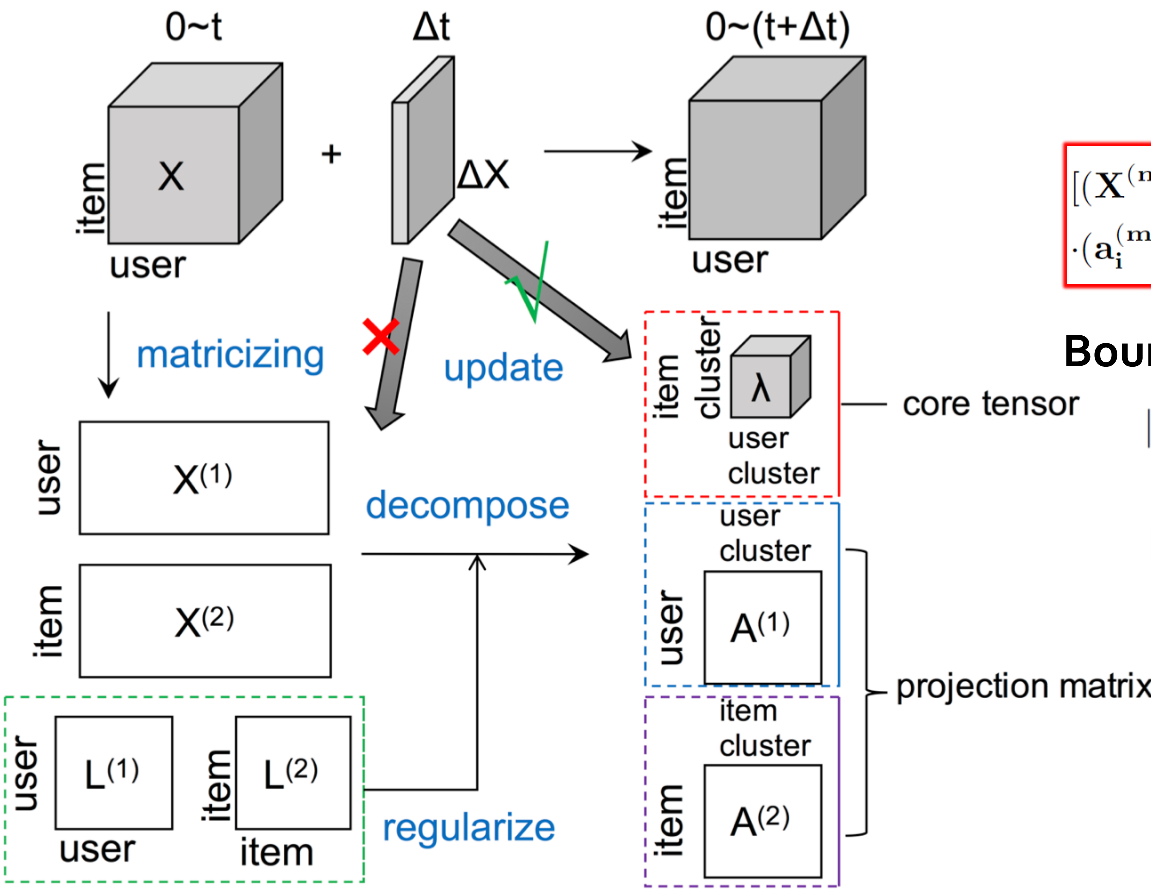
Q4-1: Who-@-whom prediction: High complexity!



M4-1: Incremental tensor decomposition: Approximation based on tensor perturbation theory. We proved bound guarantees of errors.

S4-1: RMSE reduced from **1.120** to **0.894** (**relatively -20.2%**); running time reduced from **25 hours** to **51 minutes**. KDD'14: cited by **28**.

S4-1: Flexible Evolutionary Multifaceted Analysis



Tensor perturbation theory:

$$
\boxed{\begin{aligned}&[(\mathbf{X}^{(\mathbf{m})}+\Delta\mathbf{X}^{(\mathbf{m})})(\mathbf{X}^{(\mathbf{m})}+\Delta\mathbf{X}^{(\mathbf{m})})^{\top}+\mu^{(m)}\mathbf{L}^{(\mathbf{m})}]\\&\cdot(\mathbf{a}_{\mathbf{i}}^{(\mathbf{m})}+\Delta\mathbf{a}_{\mathbf{i}}^{(\mathbf{m})})=(\lambda_i^{(m)}+\Delta\lambda_i^{(m)})(\mathbf{a}_{\mathbf{i}}^{(\mathbf{m})}+\Delta\mathbf{a}_{\mathbf{i}}^{(\mathbf{m})})\end{aligned}}
$$

Bounds (guarantee for approximation):

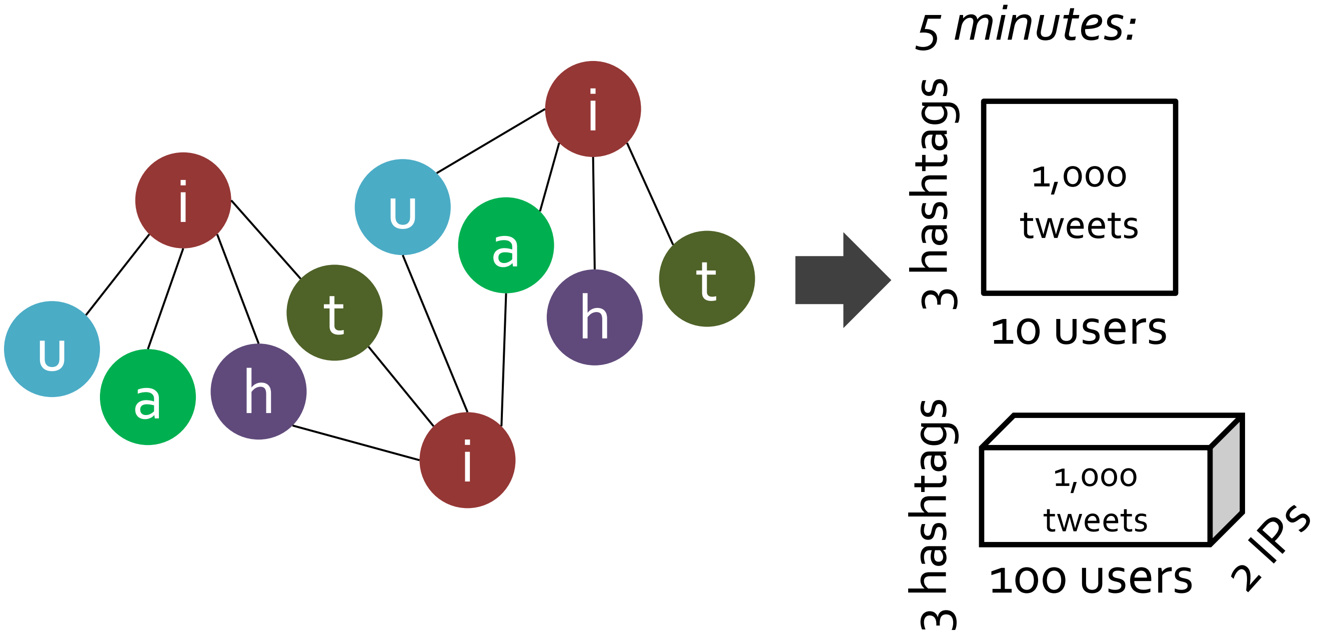
$$
|\Delta\lambda_i^{(m)}|\leq 2(\lambda^{\max}_{\mathbf{X}^{(\mathbf{m})\top}\mathbf{X}^{(\mathbf{m})}})^{\frac{1}{2}}\|\Delta\mathbf{X}^{(\mathbf{m})}\|_2
$$

$$
|\Delta\mathbf{a}_{\mathbf{i}}^{(\mathbf{m})}|\leq 2\|\Delta\mathbf{X}^{(\mathbf{m})}\|_2\sum_{j\neq i}\frac{(\lambda^{\max}_{\mathbf{X}^{(\mathbf{m})\top}\mathbf{X}^{(\mathbf{m})}})^{\frac{1}{2}}}{|\lambda_i^{(m)}-\lambda_j^{(m)}|}
$$

Q4: Knowledge from Spatiotemporal Information
M4: Tensor Methods for Modeling Multiple Dimensions

Spatiotemporal contexts

Q4-2: Spam detection: Evaluating suspiciousness?

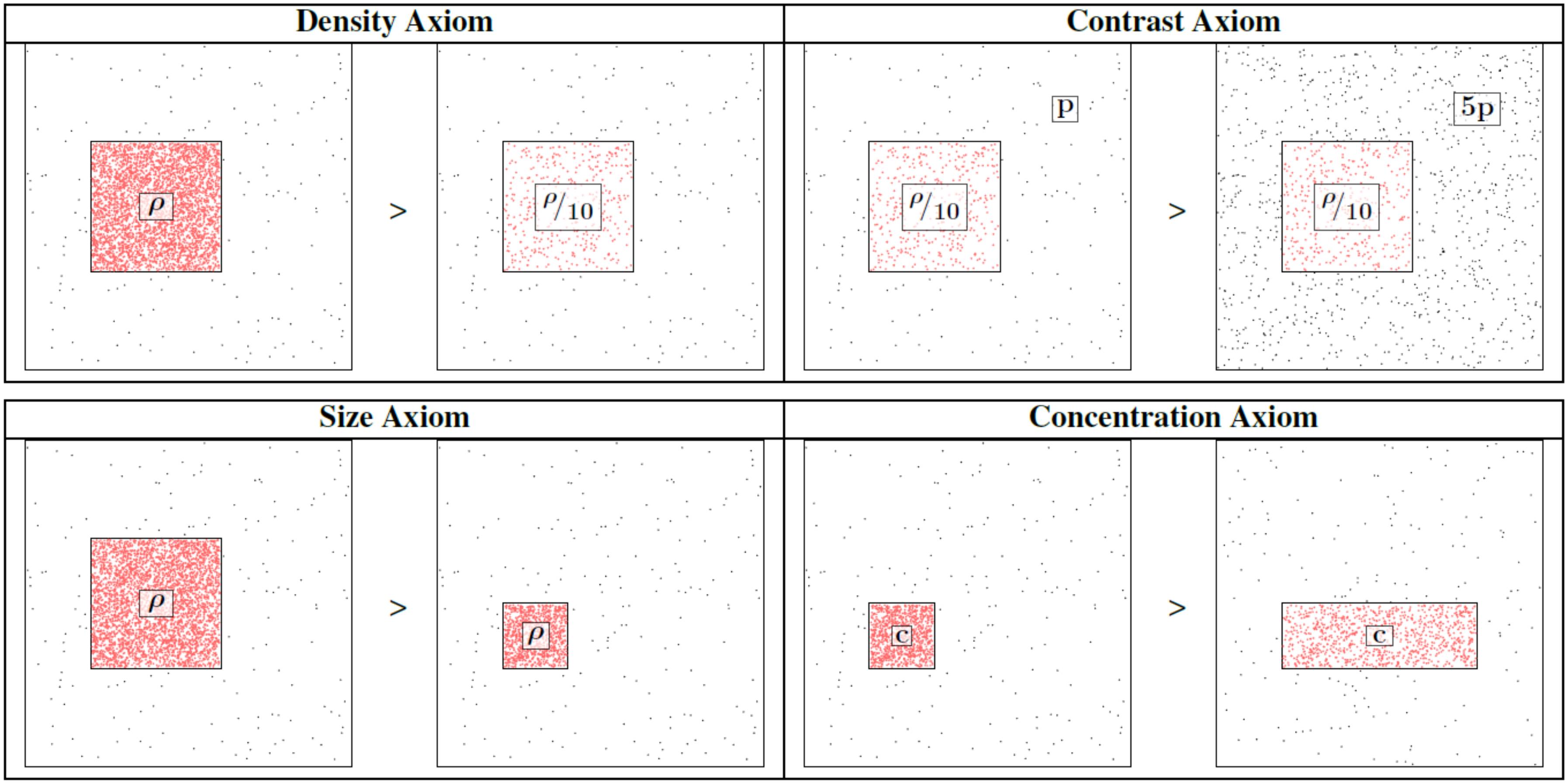


M4-2: Proposed a principled suspiciousness metric.

S4-2: Detected ~6M hijacking tweets of 3 hashtags by ~600 users from ~300 IP addresses in ~40 days.

ICDM'15: cited by **42**; TKDE'16: cited by **8**.

S4-2: Evaluating Suspiciousness across Dimensions



Q5: Knowledge from Behavioral Content

From Words, Topics, to Networks

"Modeling Complex Behaviors in Social Media", July 2015. 🎓

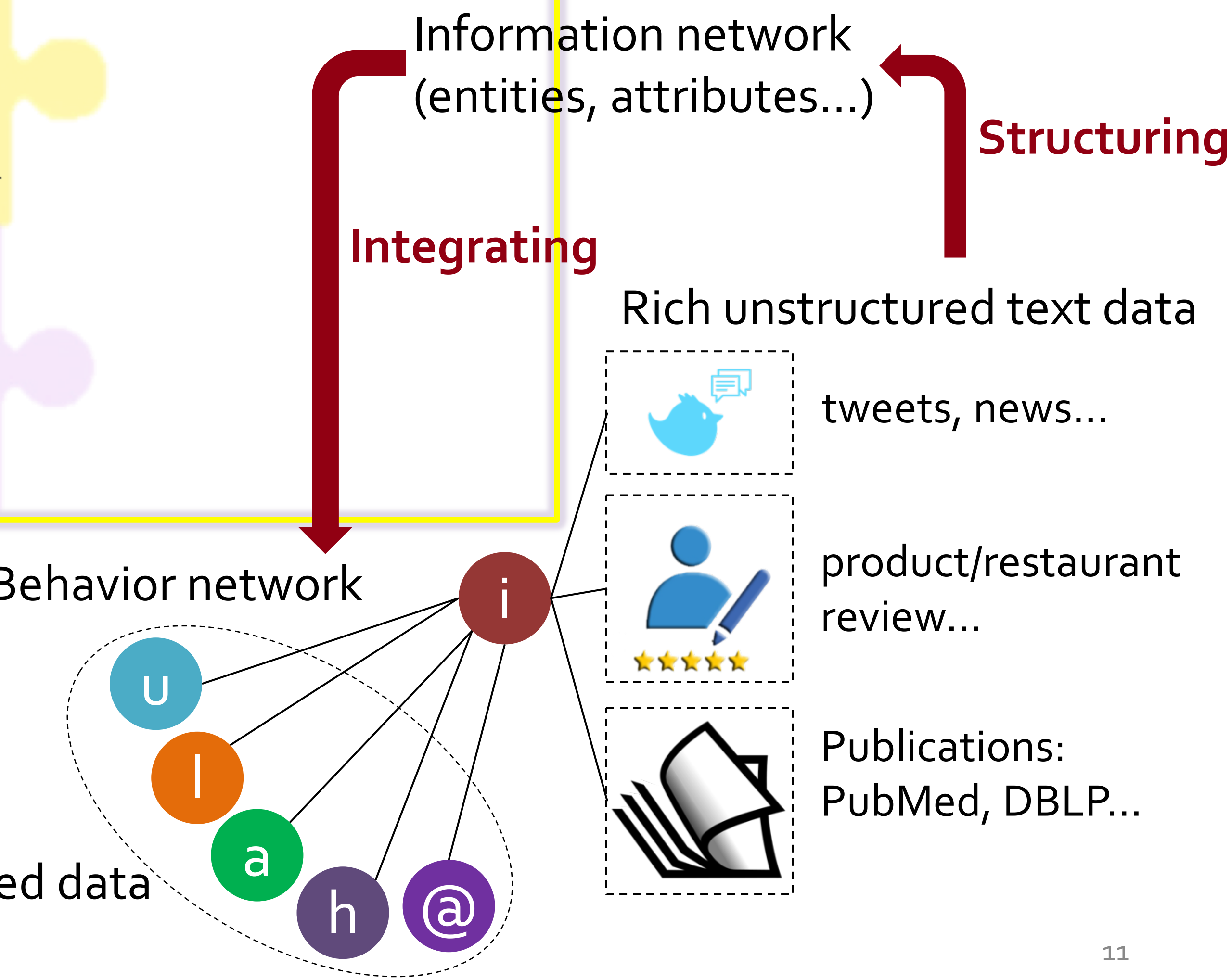
清華大學
Tsinghua University

Q5: Knowledge from Behavioral Content

From Words, Topics, to Networks

"Modeling Complex Behaviors in Social Media", July 2015.

Tsinghua University



Q5’: Entity and Attribute Discovery

Given full **text** of all the Data Science publications

Q’5-1. Who has studied the biggest number of datasets of large scale?

Q’5-2. Who study truly big data, and who always claim their work is on big data but their datasets are not big at all?

Q’5-3. For a dataset and a problem, who are the experts? How can we organize a team to solve the problem?

Q5': Entity and Attribute Discovery

Given full **text** of all the Data Science publications

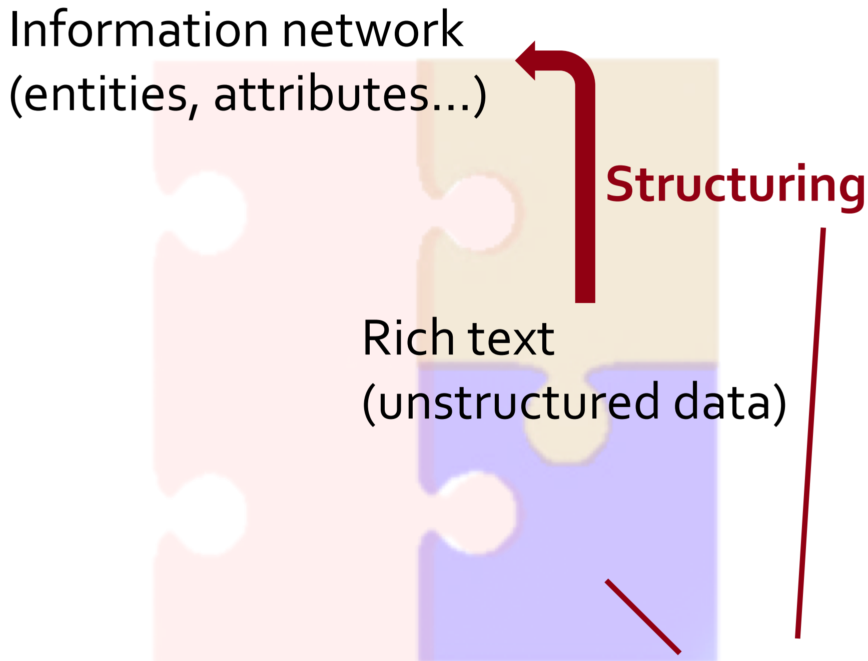
Q'5-1. Who has studied the biggest number of datasets of large scale?

Q'5-2. Who study truly big data, and who always claim their work is on big data but their datasets are not big at all?

Q'5-3. For a dataset and a problem, who are the experts? How can we organize a team to solve the problem?

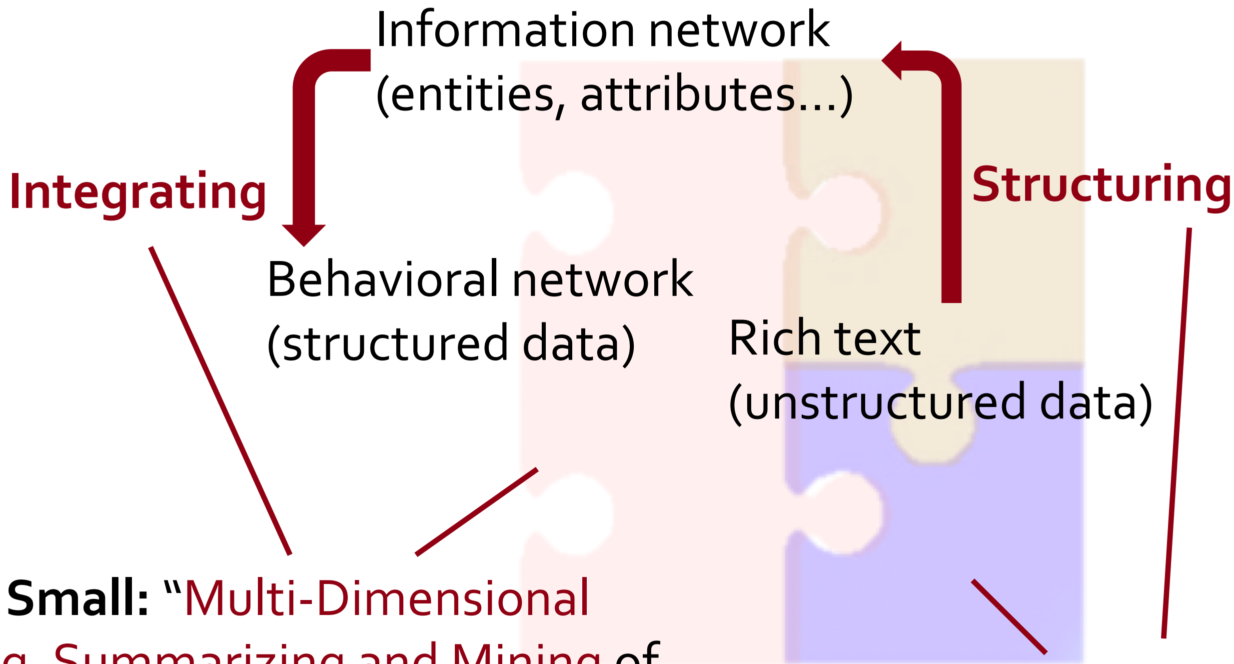
Sorry, I don't have answers now… But…

S5: Multiple Proposal Writing and Papers



1. **NSF III: Medium:** Collaborative: "StructNet: Constructing and Mining Structure-Rich Information Networks for Scientific Research". (Funded 2017)
2. **KDD'17:** "Meta Pattern-Driven Attribute Discovery from Massive Text Corpora". (Accepted)

S5: Multiple Proposal Writing and Papers



3. **NSF III: Small:** "Multi-Dimensional Structuring, Summarizing and Mining of Social Media Data", NSF IIS 16-18481. Jiawei Han, PI. (Submitted Nov 15, funded Aug 16)

4. **KDD'16:** "CatchTartan: Representing and Summarizing Dynamic Multicontextual Behaviors". **Oral** (Acc. = 8.9%).

1. **NSF III: Medium:** Collaborative: "StructNet: Constructing and Mining Structure-Rich Information Networks for Scientific Research". (Funded 2017)

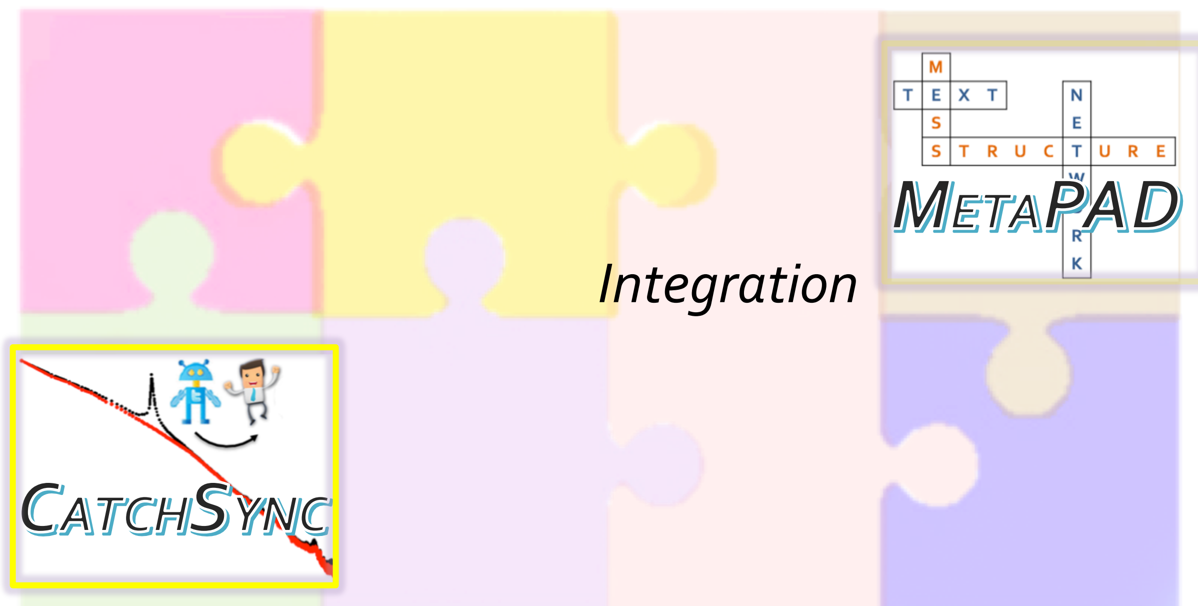
2. **KDD'17:** "Meta Pattern-Driven Attribute Discovery from Massive Text Corpora". (Accepted)

Outline

Social contexts | Spatiotemporal contexts | Behavioral content

Intelligence:
Behavior prediction and recommendation

Trustworthiness:
Suspicious behavior detection



CatchSync: Catching Synchronized Behavior in Large Directed Graphs

Joint work with Peng Cui, Shiqiang Yang (Tsinghua),

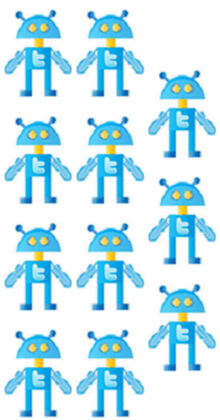
Alex Beutel, and Christos Faloutsos (CMU)

ACM SIGKDD 2014 Best Paper Finalist

(among **151** accepted research papers, **1,036** submissions)

Q3: Catching Zombie Followers

Q3: Catching Zombie Followers



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

engineers

product managers

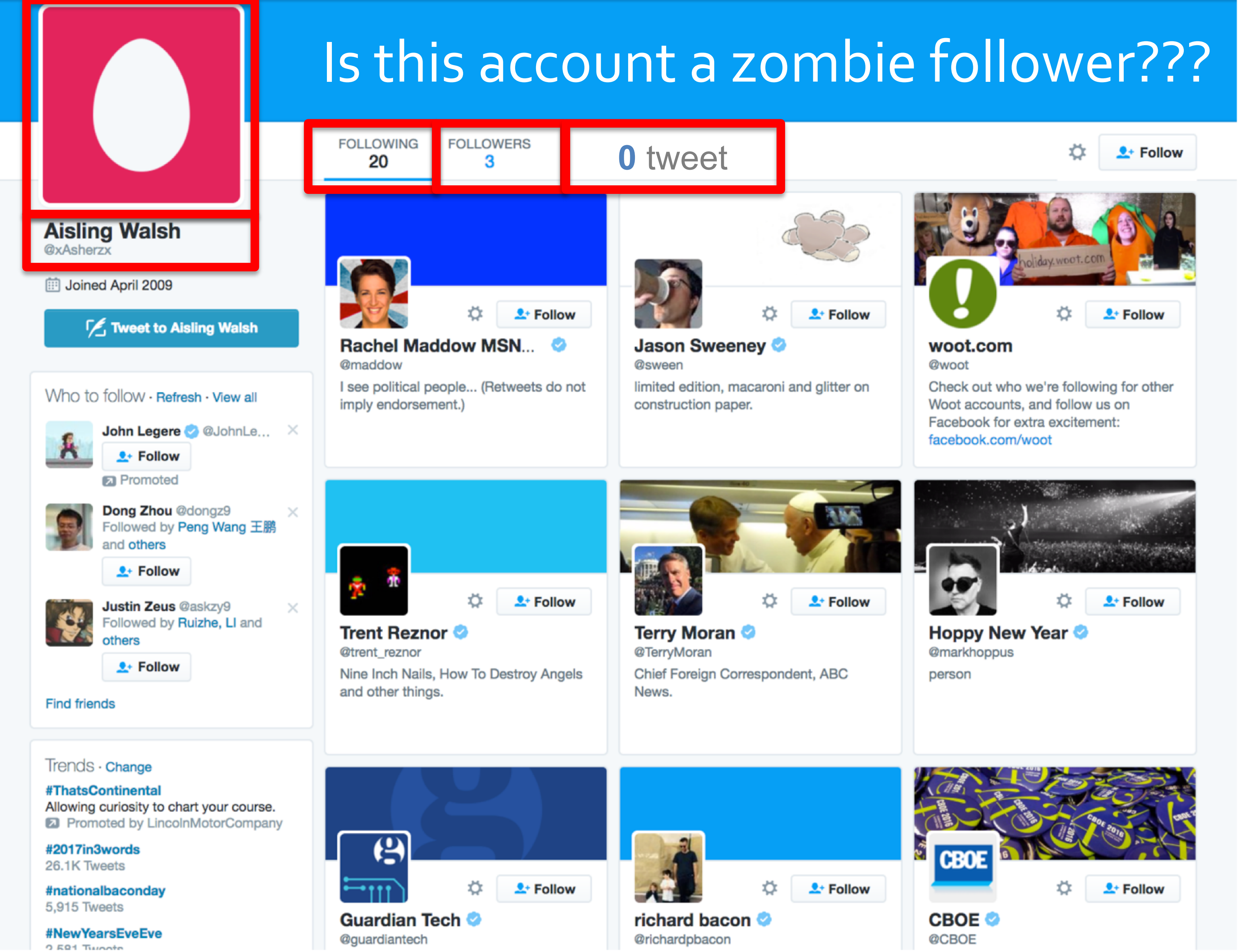
Knowledge from manual inspection: #followees, #followers, #tweets, #hashtags, #urls…

Learning models (classifiers)

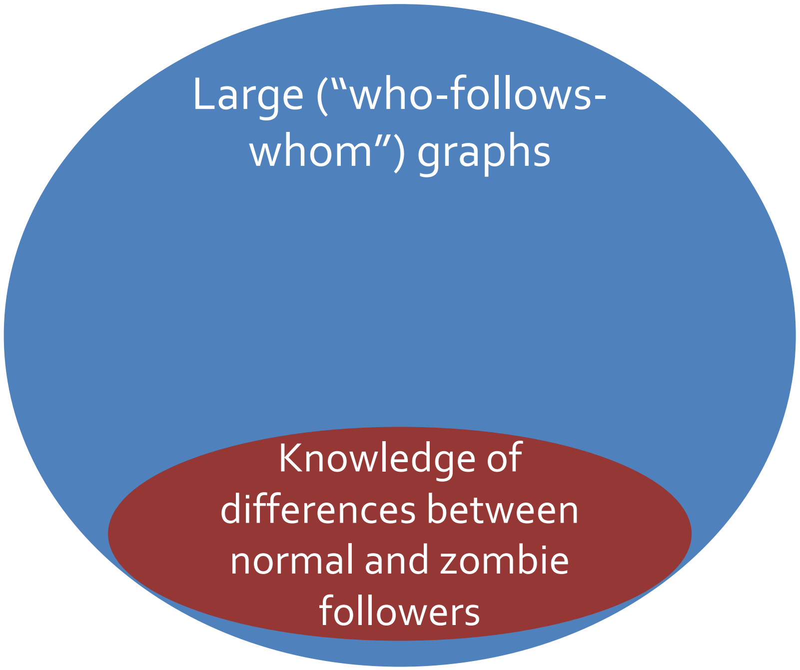
Poor accuracy
(**serious complaints** from users)

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???



Our Methodology (M1)



Our Methodology (M1)

Large ("who-follows-whom") graphs

Knowledge of differences between normal and zombie followers

Behavioral theory
(Think what they have to do, not what they can do!)

Consistently Connecting to Customers

Our Methodology (M1)

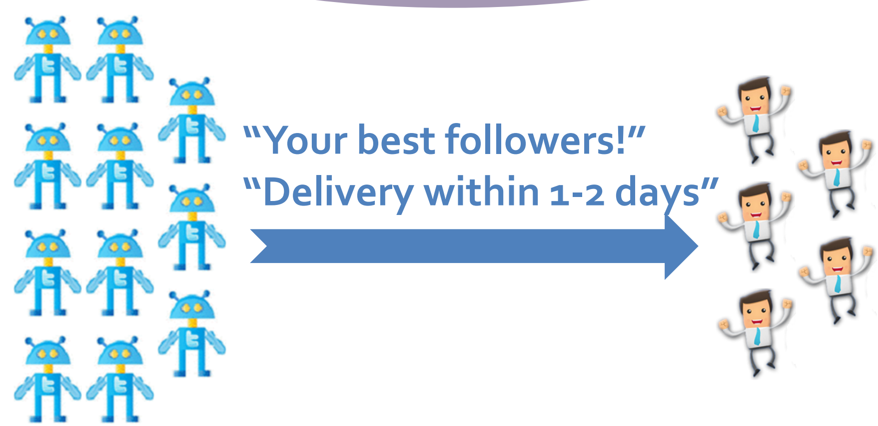
Large ("who-follows-whom") graphs

Knowledge of differences between normal and zombie followers

Behavioral theory (Think what they have to do, not what they can do!)

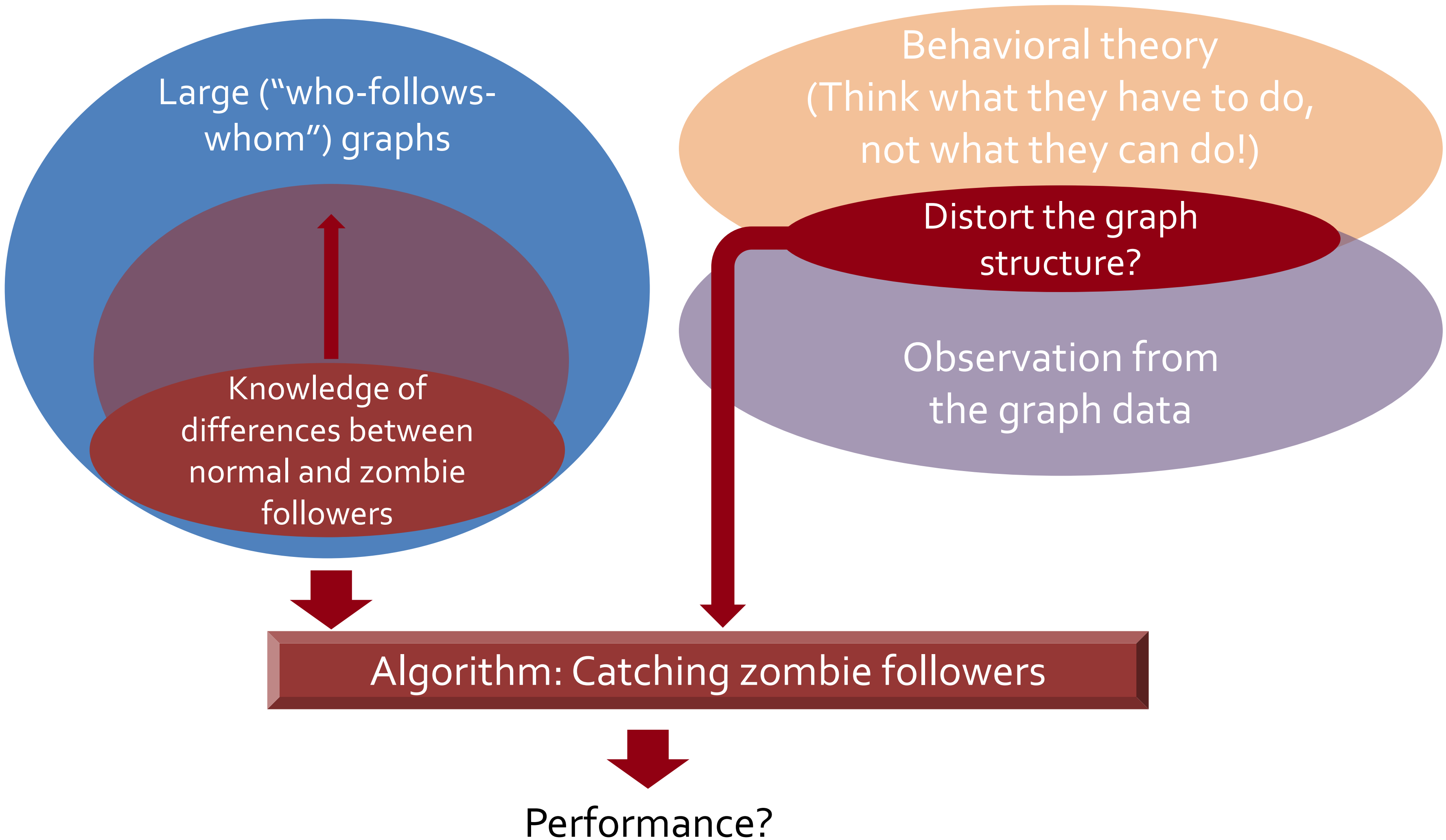
Distort the graph structure?

Observation from the graph data



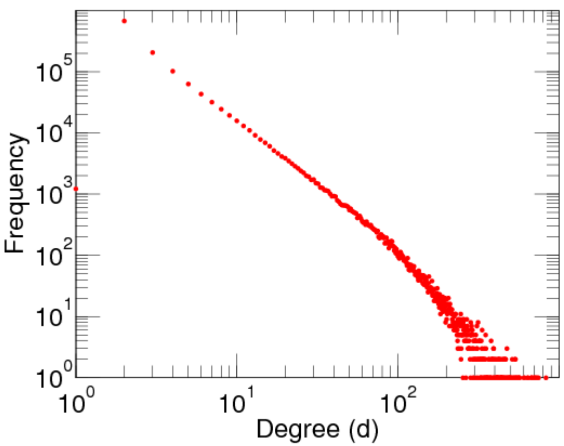
Consistently Connecting to Customers

Our Methodology (M1)

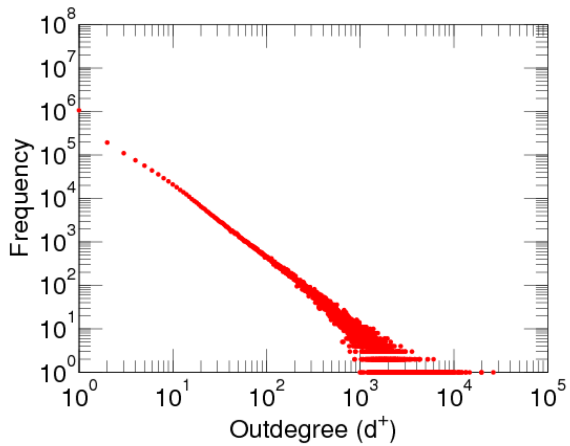


Out-Degree Distributions: Power Law Expected

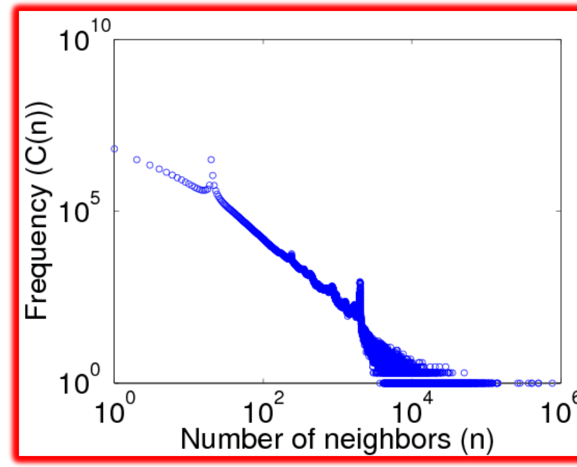
DBLP
Author-publication



Flickr
User-user



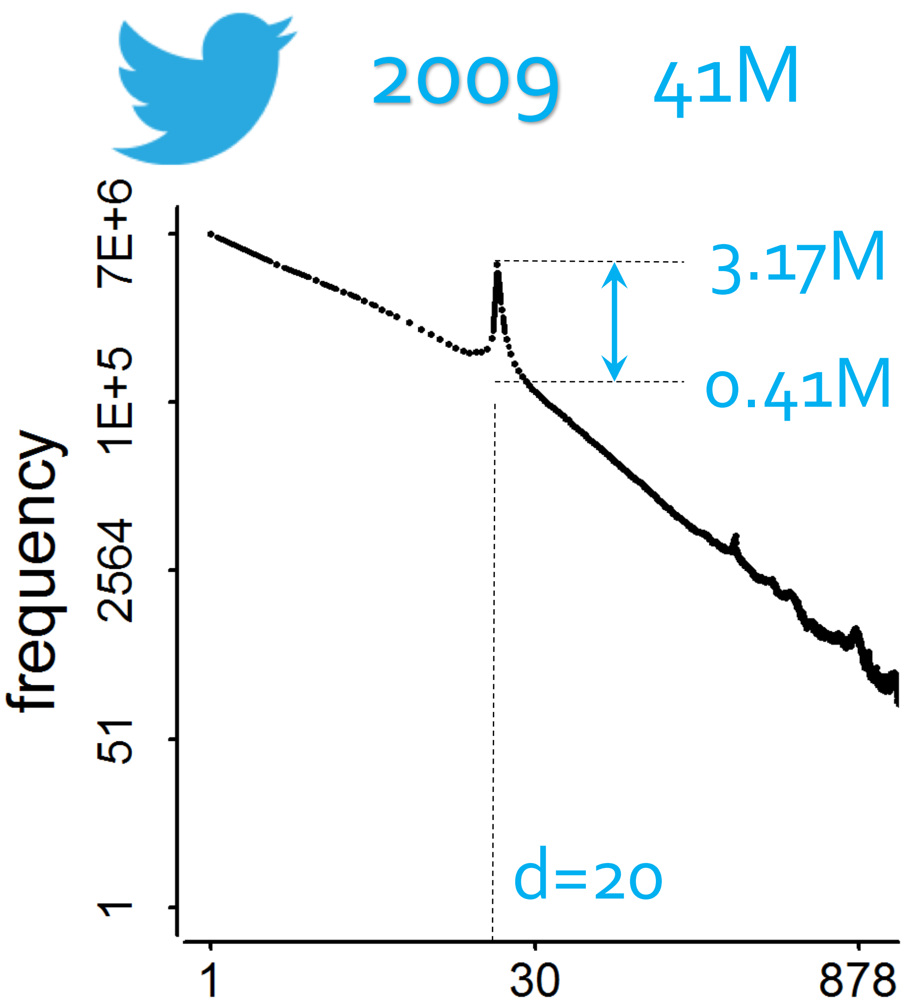
Twitter
Who-follows-whom



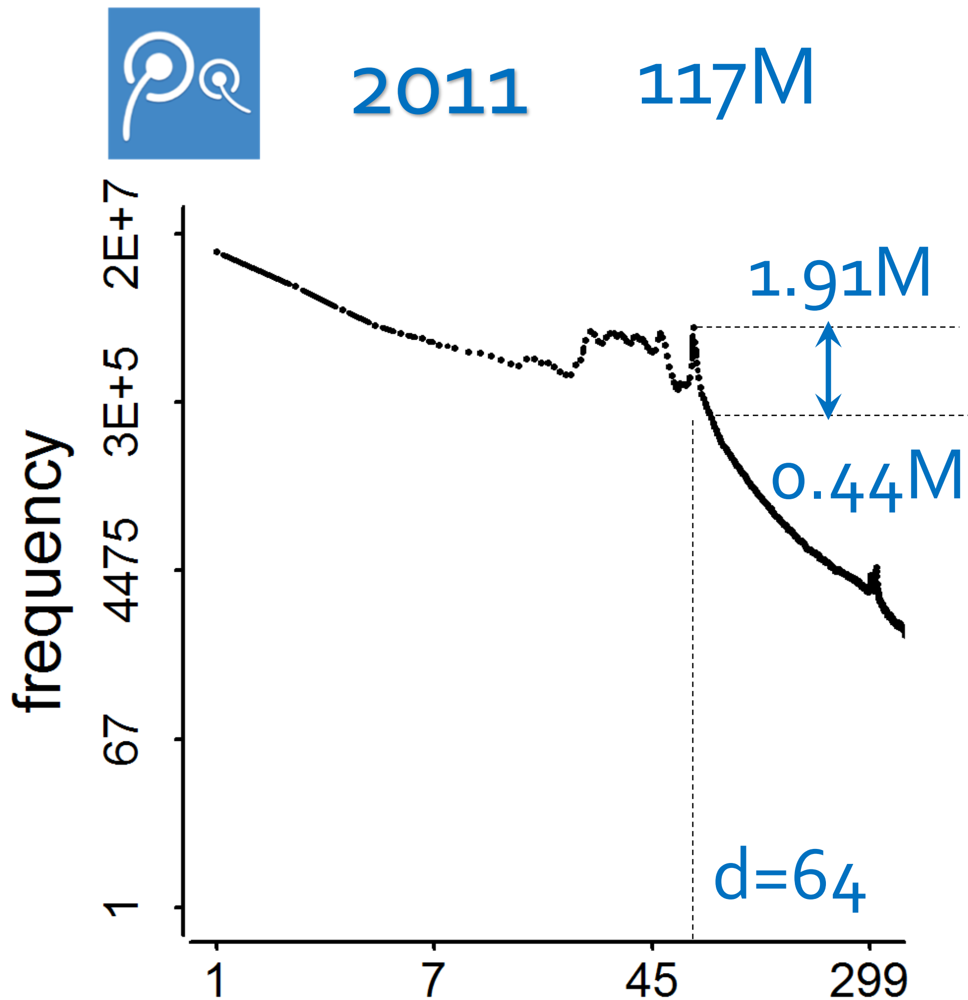
[konect.uni-koblenz.de/networks/]

Power-law distributions in networks [Faloutsos et al. SIGCOMM'99; Chung et al. PNAS'02]

Spikes!

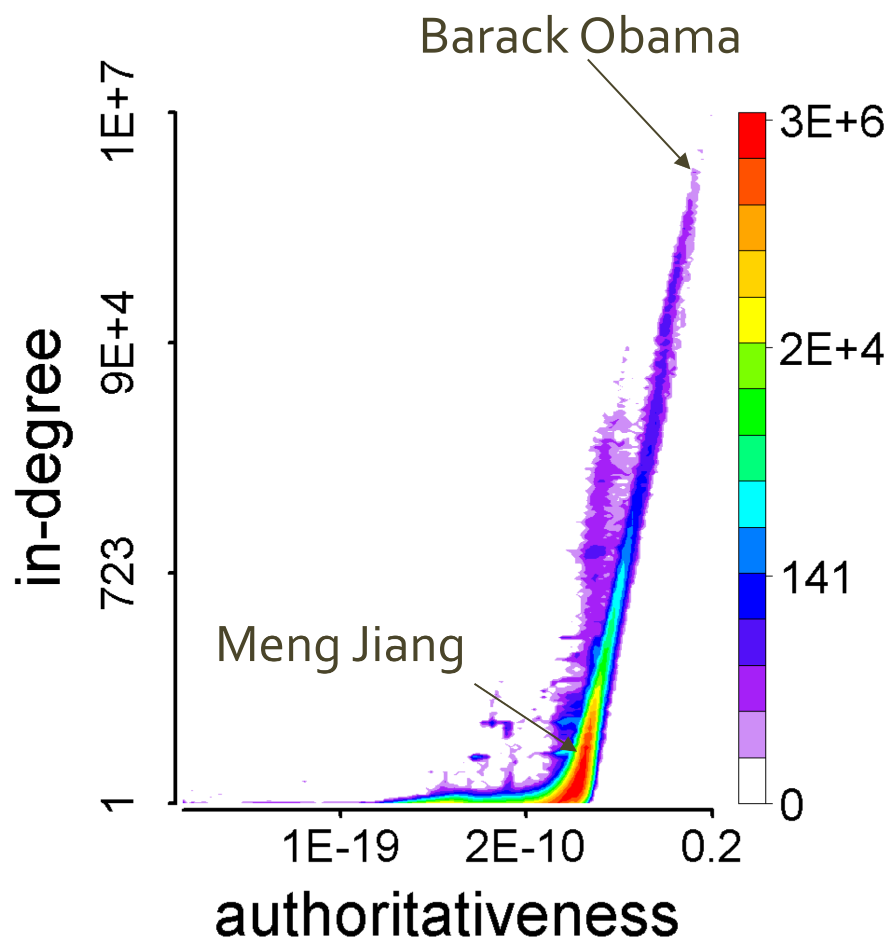


2009
41M
frequency
7E+6
1E+5
2564
51
1
3.17M
0.41M
d=20
1
30
878
out-degree



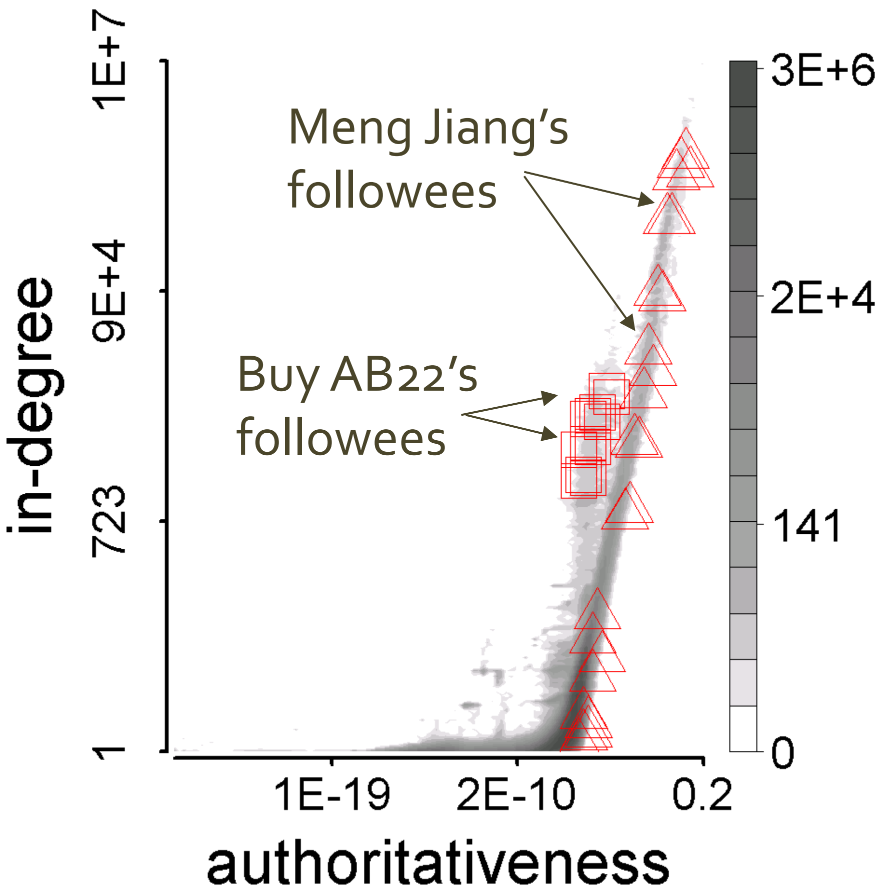
2011
117M
frequency
2E+7
3E+5
4475
67
1
1.91M
0.44M
d=64
1
7
45
299
out-degree

How We/They Connect to Our/Their Followees

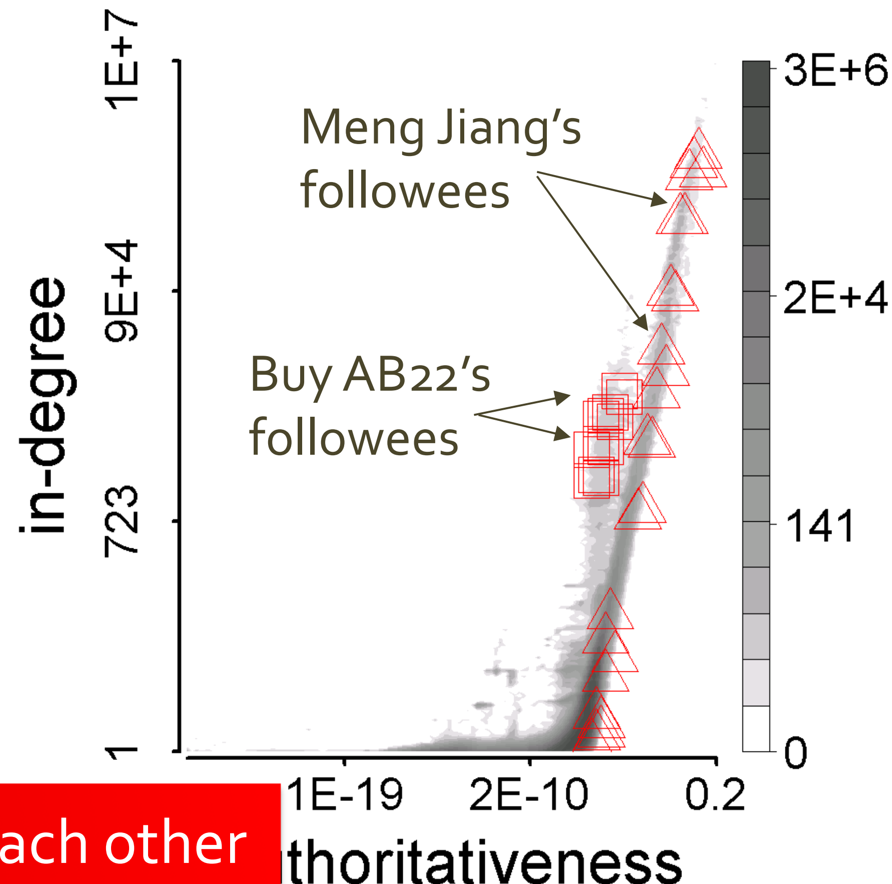


The HITS algorithm. Kleinberg. "Authoritative sources in a hyperlinked environment." JACM'99.

How We/They Connect to Our/Their Followees



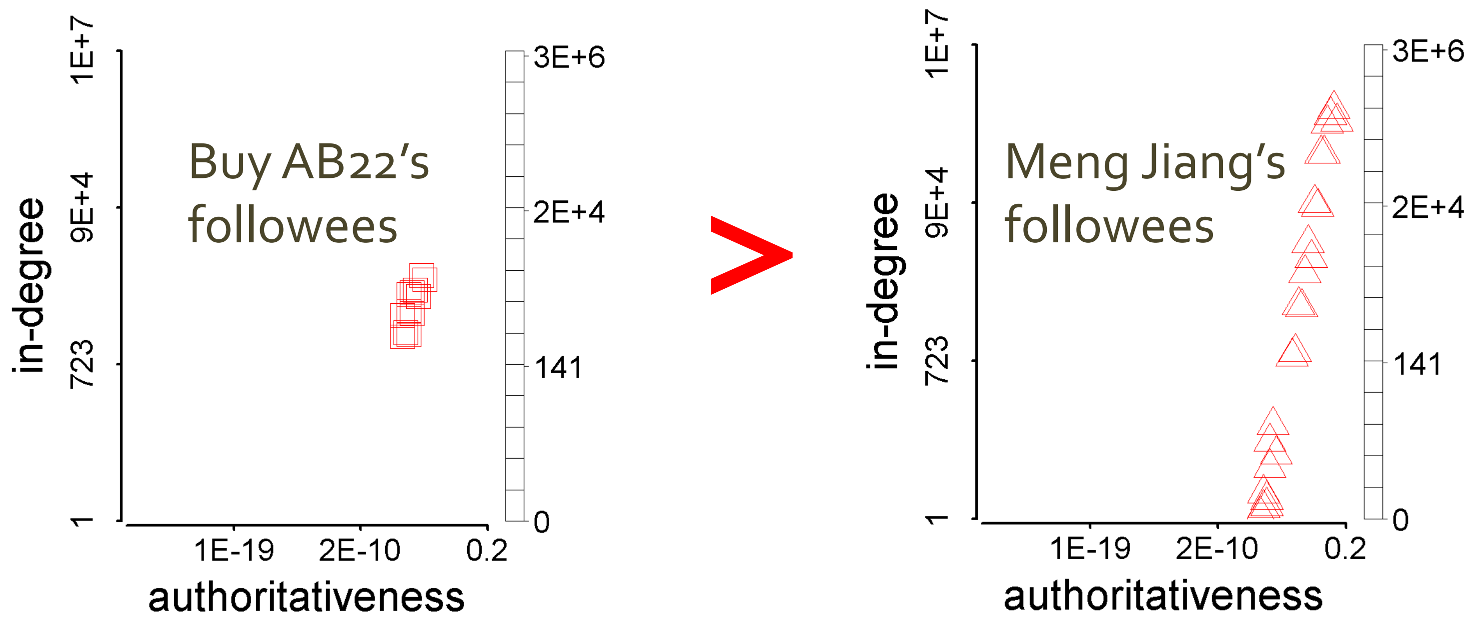
How We/They Connect to Our/Their Followees



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

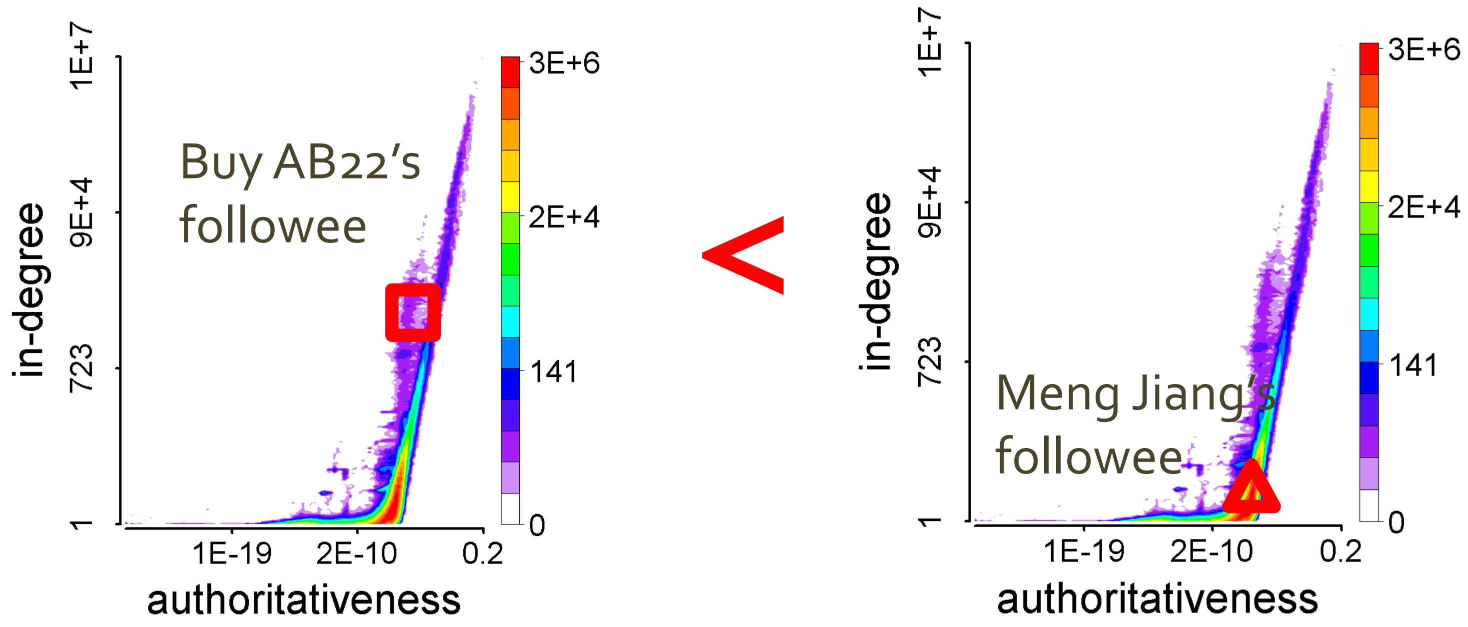
Definition: Synchronicity

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



Definition: Normality

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



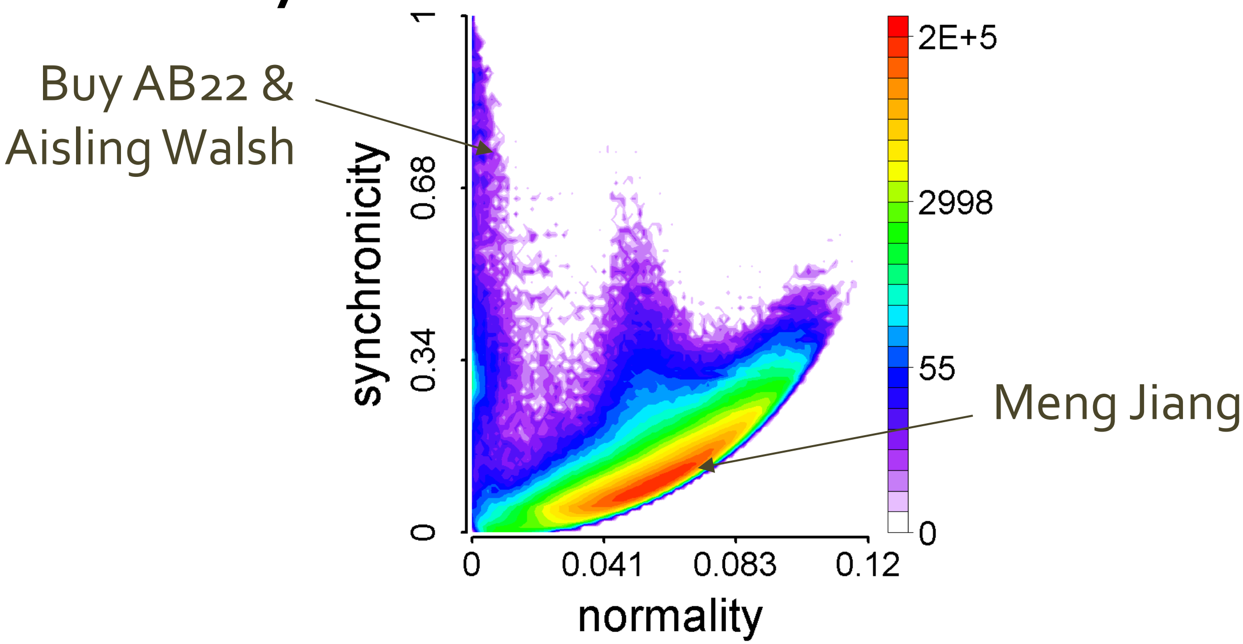
When is the Synchronicity Too High?

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

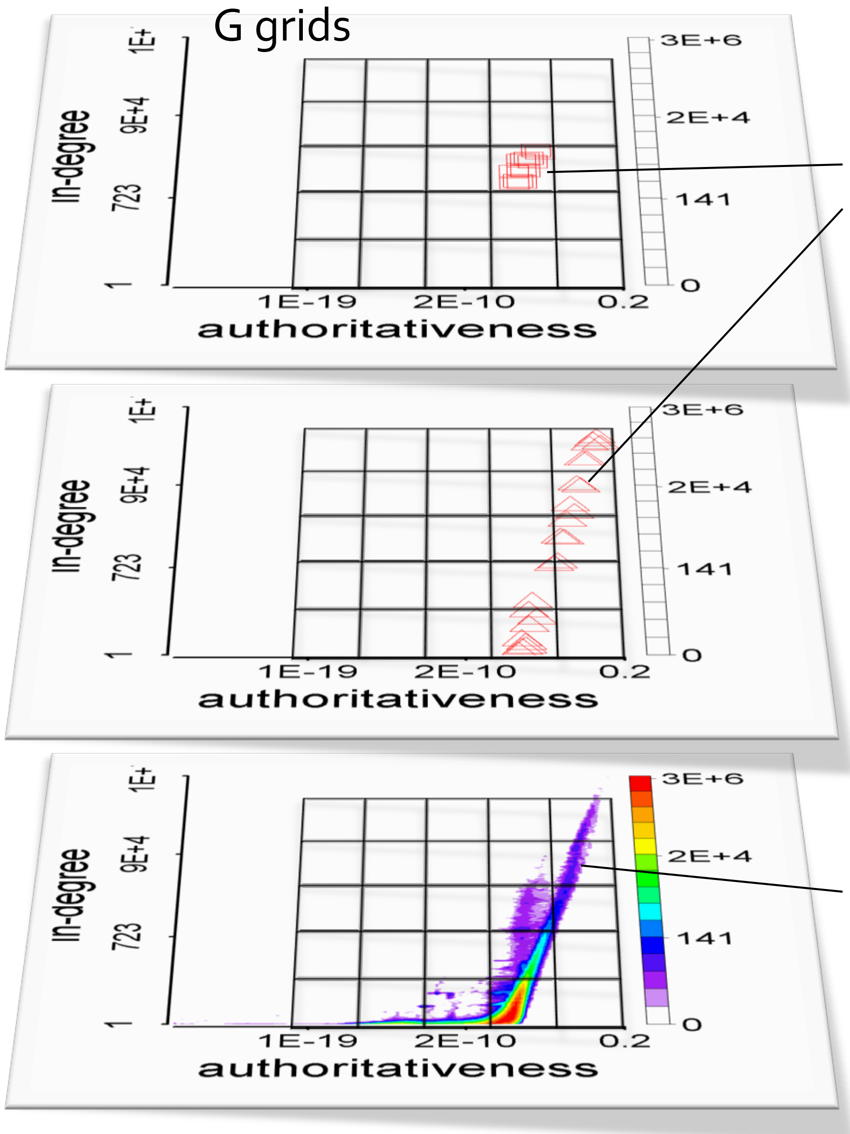
Theorem:

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

Our CatchSync:



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

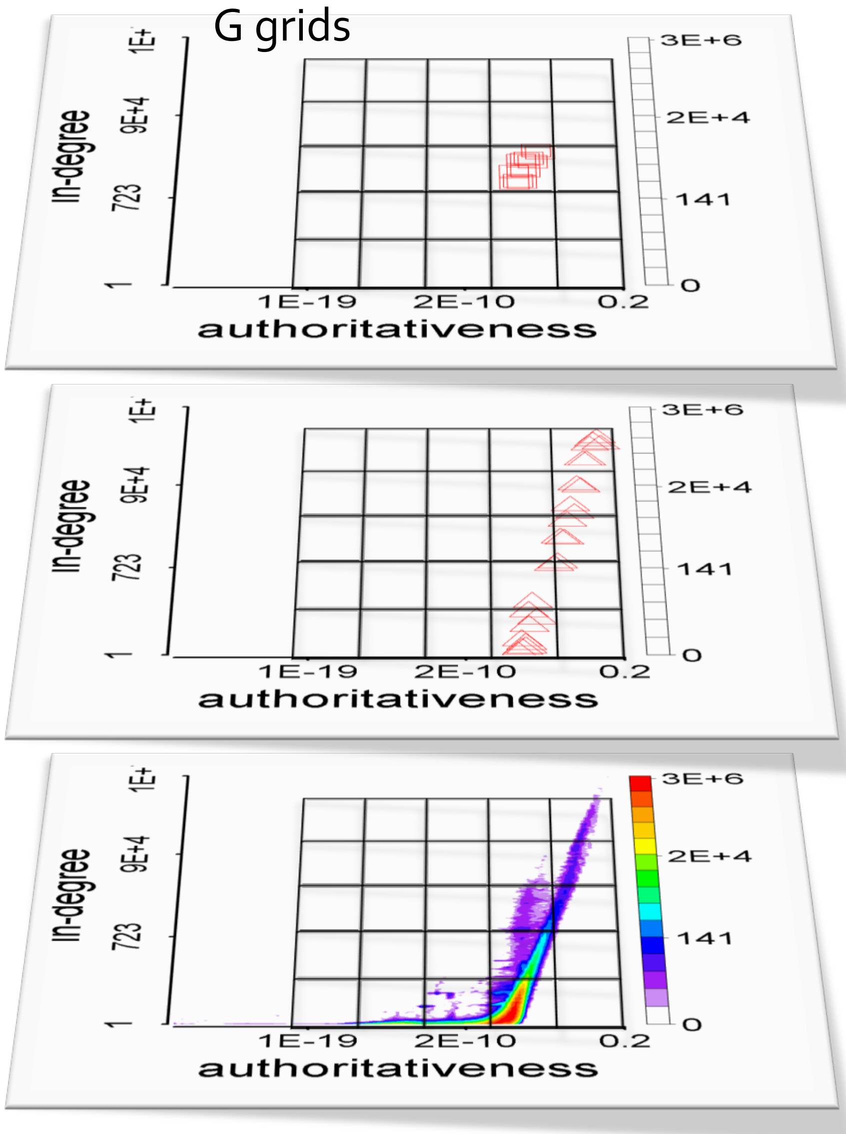
$$s_{min} = \sum (fp_g/F)(fp_g/F) = \sum f_g^2$$

where

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

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

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:

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

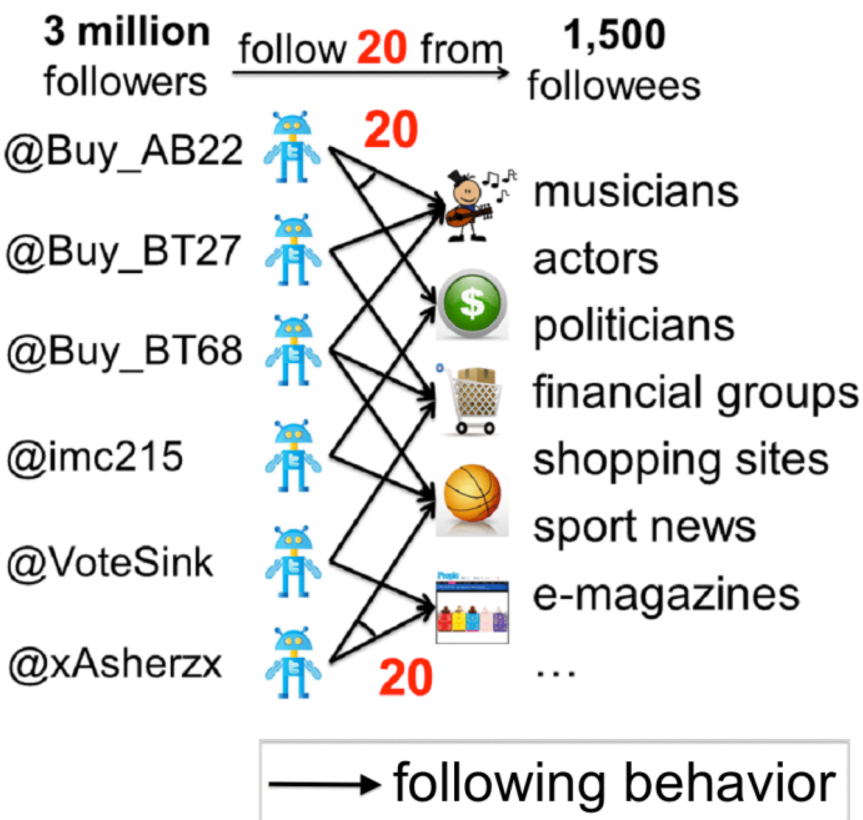
$$\begin{cases} 2 + \lambda G + \mu = 0 & \xleftarrow{\;\sum\;} \\ 2n + \lambda + \mu s_b = 0 & \xleftarrow{\;\times b_g \sum\;} \\ 2 s_{min} + \lambda + \mu n = 0 & \xleftarrow{\;\times f_g \sum\;} \end{cases}$$

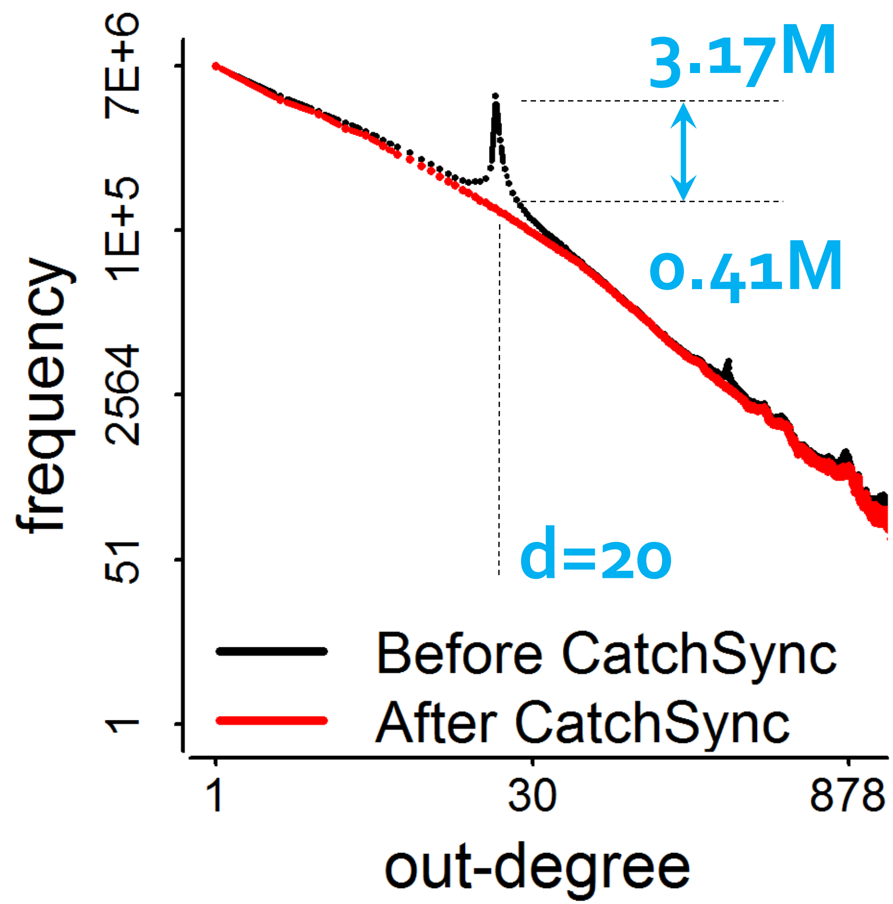
where $s_b = \sum b_g^2$.

Therefore,

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

The Distribution was Recovered!





Impact

- Cited by **76**; WWW'14: cited by **21**. TKDD'16: cited by **21**. Synchronized behavior in cyber attacks.
  - ACM SIGSAC Conference on **Computer and Communications Security** (CCS), 2015 : two papers
  - **Network and Distributed System Security** Symposium (NDSS), 2016: one paper
  - The 31st Annual **Computer Security Applications** Conference (ACSAC), 2015: one paper
  - IEEE Transactions on **Information Forensics and Security** (TIFS), 2016: one paper
  - International Journal of **Digital Crime and Forensics** (IJDCF), 2016: one paper
  - …

Impact

- Cited by **76**; WWW'14: cited by **21**. TKDD'16: cited by **21**. Synchronized behavior in cyber attacks.
- Taught in
  - CMU 15-826: Multimedia Databases and Data Mining
  - UMich EECS 598: Graph Mining and Exploration at Scale
  - ASONAM'16 Tutorial: "Identifying Malicious Actors on Social Media" by S. Kumar, F. Spezzano, V.S. Subrahmanian
- Endless games! First proposed **Camouflage** in PAKDD'14.
  - Cited by **47**.
  - Cited by *KDD'16 Best Research Paper*: the authors (B. Hooi *et al.*) provided theoretical bounds to prevent the camouflage.

References

Breese, Heckerman, and Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In the 4th conference on Uncertainty in artificial intelligence (UAI), 1998.

Getoor and Sahami. Using probabilistic relational models for collaborative filtering. In Workshop on Web Usage Analysis and User Profiling (WEBKDD), 1999.

Herlocker, Konstan, and Riedl. Explaining collaborative filtering recommendations. ACM Conference on Computer Supported Cooperative Work (CSCW), 2000.

Herlocker, Konstan, Terveen, and Riedl. Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems (TOIS) 22, no.1, 2004.

Yu and Ji. Unsupervised Person Slot Filling based on Graph Mining. ACL, 2016.

Hu, Koren, and Volinsky. Collaborative filtering for implicit feedback datasets. In International Conference on Data Mining (ICDM), pages 263–272, 2008.

Koren. Factorization meets the neighborhood: A multifaceted collaborative filtering model. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2008.

Koren, Bell, and Volinsky. Matrix factorization techniques for recommender systems. Computer, 2009.

Liu and Yang. Eigenrank: A ranking-oriented approach to collaborative filtering. International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2008.

Liu, Zhao, and Yang. Probabilistic latent preference analysis for collaborative filtering. ACM International Conference on Information and Knowledge Management (CIKM), 2009.

Salganik, Dodds, and Watts. Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. Science, 2006.

References

Sarwar, Karypis, Konstan, and Riedl. Item-based collaborative filtering recommendation algorithms. International conference on World Wide Web (WWW), 2001.

Burke. Hybrid recommender systems: Survey and experiments. User modeling and user-adapted interaction, 2002.

Ma, Zhou, Lyu, and King. Improving recommender systems by incorporating social contextual information. ACM Transactions on Information Systems (TOIS), 2011.

Ma. King, and Lyu. Learning to recommend with explicit and implicit social relations. ACM Transactions on Intelligent Systems and Technology (TIST), 2011.

Tang, Hu, and Liu. Social recommendation: a review. Social Network Analysis and Mining, 2013.

Han, Zhang, Ghalwash, Vucetic, and Obradovic. Joint Learning of Representation and Structure for Sparse Regression on Graphs. SIAM International Conference on Data Mining (SDM), 2016.

Egele, Stringhini, Kruegel, and Vigna. COMPA: Detecting Compromised Accounts on Social Networks. The Network and Distributed System Security Symposium (NDSS), 2013.

Yang, Wilson, Wang, Gao, Zhao, and Dai. Uncovering social network sybils in the wild. ACM Transactions on Knowledge Discovery from Data (TKDD), 2014.

Viswanath, Bashir, Crovella, Guha, Gummadi, Krishnamurthy, and Mislove. Towards detecting anomalous user behavior in online social networks. USENIX Security Symposium (USENIX Security), 2014.

Faloutsos, Faloutsos, and Faloutsos. On power-law relationships of the internet topology. ACM SIGCOMM computer communication review, 1999.

Chung, and Lu. The average distances in random graphs with given expected degrees. Proceedings of the National Academy of Sciences, 2002.

References

Kleinberg. Authoritative sources in a hyperlinked environment. Journal of ACM, 1999.

Hooi, Song, Beutel, Shah, Shin, and Faloutsos. Fraudar: Bounding graph fraud in the face of camouflage. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2016.

Angeli, Premkumar, and Manning. Leveraging linguistic structure for open domain information extraction. Linguistics, 2015.

Schmitz, Bart, Soderland, and Etzioni. Open language learning for information extraction. Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP), 2012.

Gupta, Halevy, Wang, Whang, and Wu. Biperpedia: An ontology for search applications. Proceedings of the VLDB Endowment (VLDB), 2014.

Halevy, Noy, Sarawagi, Whang, and Yu. Discovering Structure in the Universe of Attribute Names. International Conference on World Wide Web (WWW), 2016.

Yahya, Whang, Gupta, and Halevy. ReNoun: Fact Extraction for Nominal Attributes. Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP), 2014.

Liu, Shang, Wang, Ren, and Han. Mining quality phrases from massive text corpora. ACM SIGMOD International Conference on Management of Data (SIGMOD), 2015.

Ren, El-Kishky, Wang, Tao, Voss, and Han. Clustype: Effective entity recognition and typing by relation phrase-based clustering. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2015.