

Data-Driven Behavioral Analytics: Observations, Representations and Models

Meng Jiang (UIUC)

Peng Cui (Tsinghua)

Jiawei Han (UIUC)

<http://www.meng-jiang.com/tutorial-cikm16.html>

What is Behavior?

□ **Definition.** Interactions made by **individuals** in conjunction with **themselves** or their **environment**. (*Wikipedia*)



Behavioral Analysis

- *Significance*. What can we discover from behavioral data?
 - *Ex.* Given every phone call/message between military leaders, scientists, businesspersons, find...

Observations

Who, what, where, when, why, how...
(scientific view)

Representations

Graph, network, matrix, tensor...
(mathematical view)

Models

Prediction, recommendation, anomaly detection...
(application view)

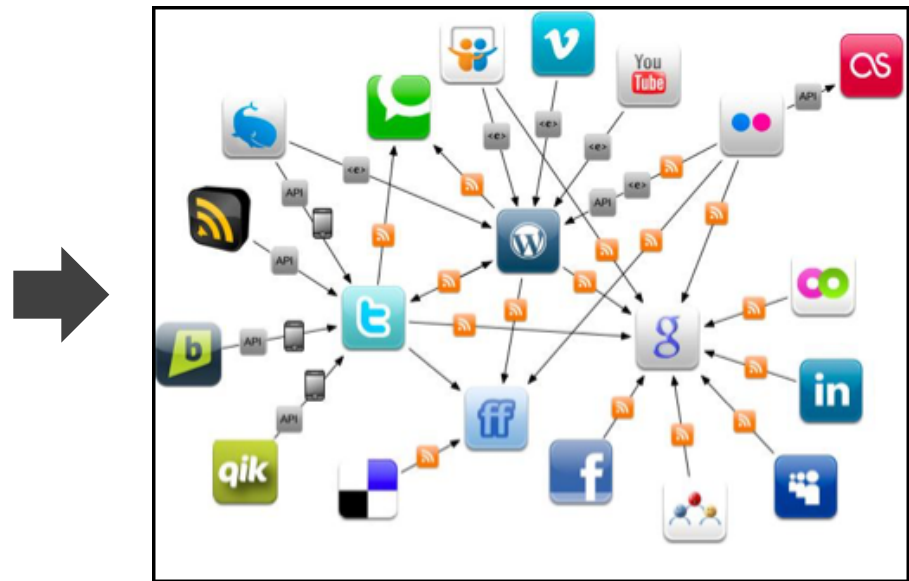
Why Behavioral Analysis Today?

□ *Today.* The human behaviors are broadly recorded in an unprecedented level. Insights of sciences and society?

Physical World



Online Applications





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

VI. Computational Models of Human Behavior



A fundamental understanding and predictive capability of human behavior dynamics from individuals to societies.

- **Enabled capabilities**
 - Predictive models supporting strategic, operational, and tactical decision making and planning
 - Real time cultural situational awareness
 - Immersive training and mission rehearsal
 - Cross cultural coalition building
- **Key research challenges:**
 - Conflicting theories
 - Data management and fusion
 - Mathematical complexity
 - Validation of models

Costly Punishment Across Human Societies

Joseph Henrich,^{1*} Michael R. McNulty,² Abigail I. Alexander-Bolger,³ Juan Carlos Cardenas,⁴ Natalie Heinrich,⁵ Carolyn L. Laxer,⁶ Frank M.

Recent behavioral experiments aimed at understanding cooperation have suggested that a willingness to punish, may be part of human psychology and, however, because most experiments have been in generalizations of these insights to the species has results from 15 diverse populations show that (i) to administer costly punishment as unequal behavior punishment varies substantially across populations with all ethnic behavior across populations. These gene-culture coevolution of human altruism and cooperation needs to explain.

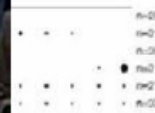
For tens of thousands of years before formal contracts, courts, and constitutions, human societies maintained important forms of cooperation in domains such as hunting, warfare, trade, and food sharing. The scale of cooperation in both contemporary and past human societies remains a puzzle for the evolutionary and social sciences, because, first, neither kin selection nor reciprocity appears to readily explain altruism in very large groups of unrelated individuals and, second, empirical assumptions of self-regarding preferences in economics and related fields appear equally ill-fitted to the facts (1). Reputation can support altruism in large groups, however, some other mechanism is needed to explain why reciprocity should be linked to prosociality rather than selfish or neutral behavior (2). Recent theoretical work



tions (13). Such experiments have even begun to probe the neural underpinnings of punishment (14, 15).

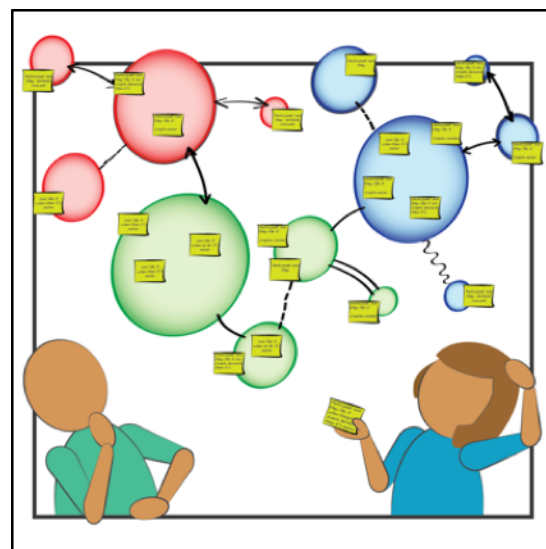
These results are important, because the role of costly punishment can explain the pieces of the puzzle of large-scale cooperation. However, like previous studies, the results used to study altruism, punishment have been conducted exclusively among university students, not know whether such findings are the peculiarities of students and/or on industrialized societies or whether indeed capturing species character. Earlier research used experimental studies in 15 diverse societies to measure giving behavior (1, 16). We found that social self-interest could not explain as in any of the 15 societies studied, found much more variation in giving than previous studies with university students. Similarly, until costly punishment is studied in more societies and university students, it is difficult to ascribe importance for explaining human behavior.

Given its importance for understanding whether costly punishment with altruistic behavior is valuable, the evolution of costly punishment in societies in which costly punishment will exhibit stronger norms and prosociality, because the



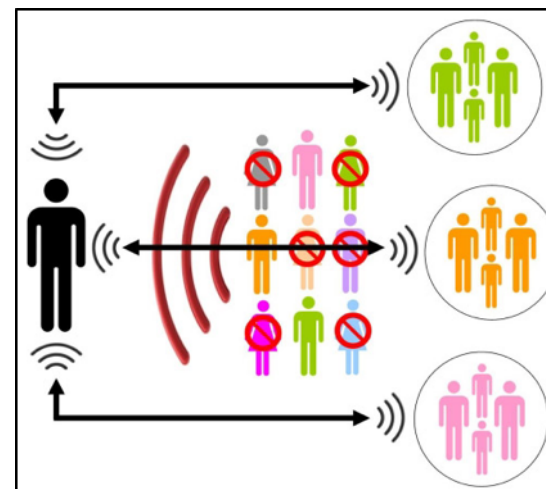
- **Measures of success**
 - Early success of simple models
 - Success of social network analysis
 - Prediction of crowd tipping points

Challenges in Behavioral Analysis



**Content
(preference)**

**Social context
(influence)**



**Behavioral
Analysis**

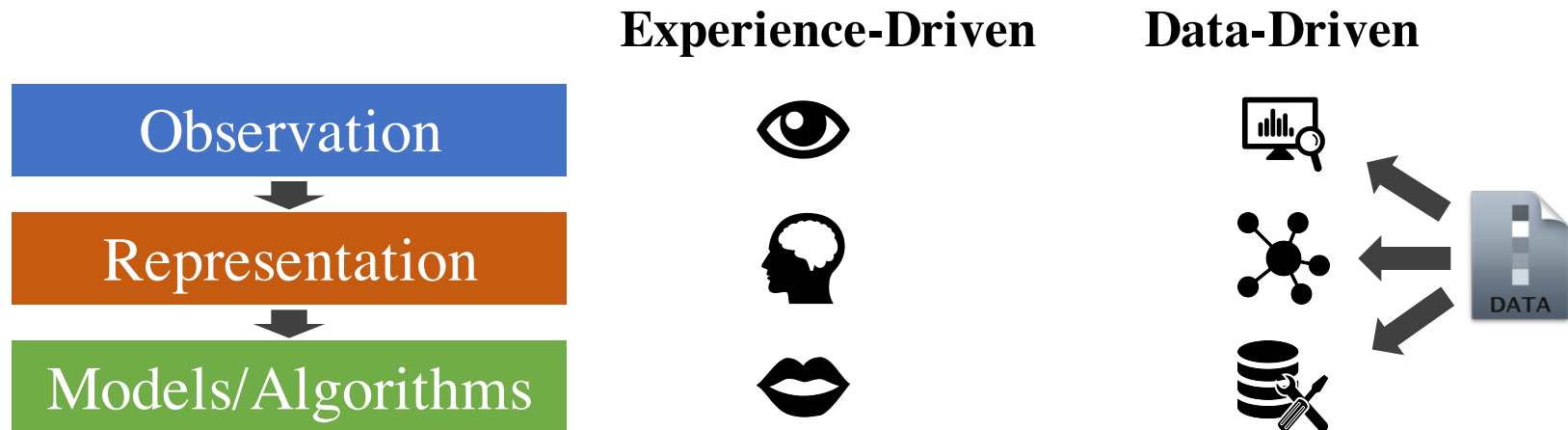
Spatiotemporal context



**Intention
(suspiciousness)**

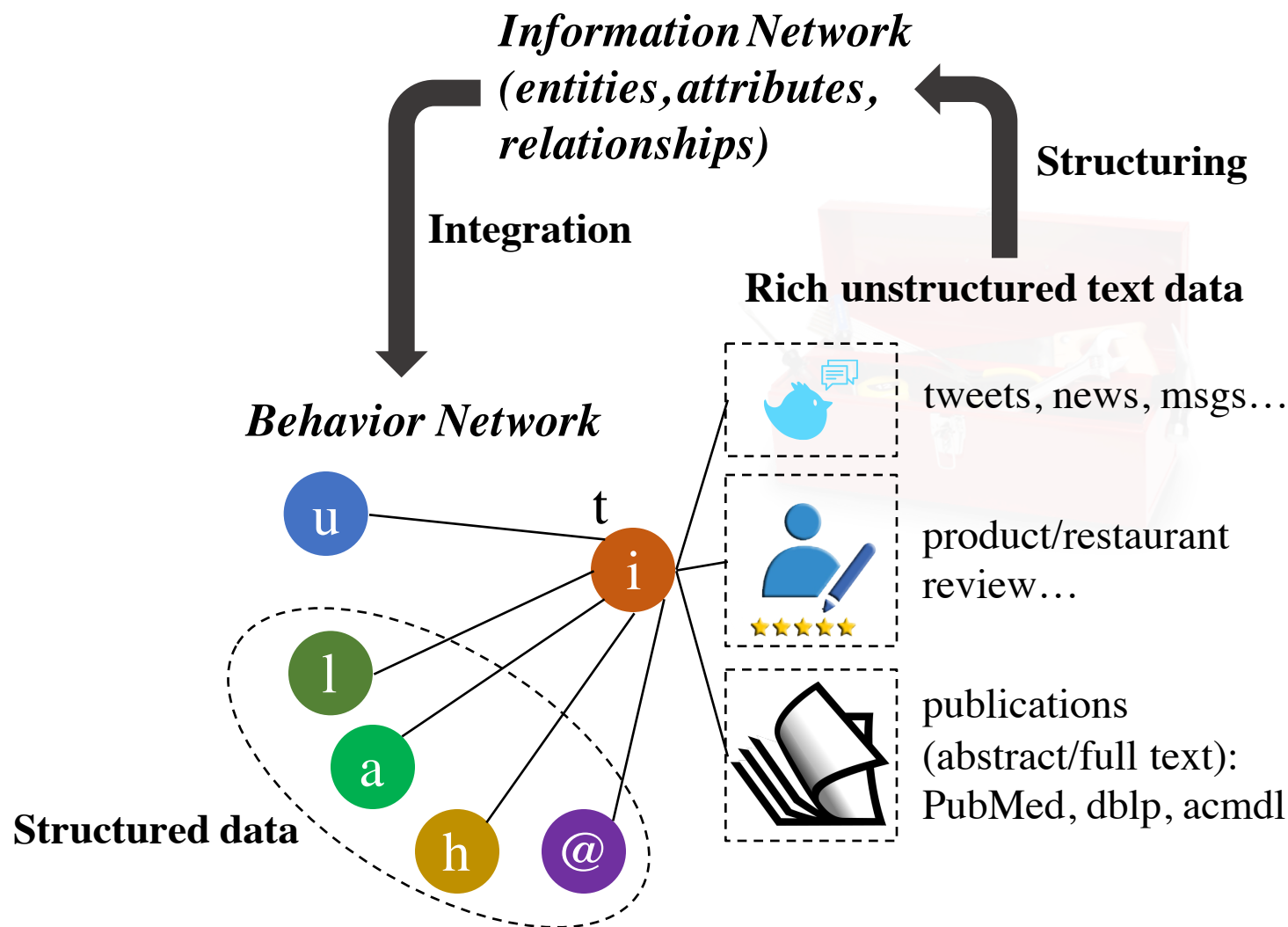
REWARDS	# TICKETS GIVEN	CONSEQUENCES	# TICKETS TAKEN AWAY
Extra Math	+5	HITTING	-3
Getting along WELL with others	+3	BULLYING	-4
Good Table Manners	+4	TEASING	-1
LOVE & RESPECT	+5	LYING	-2
Obedying the FIRST TIME	+3	THROWING A FIT	-3
Calm & Quiet in STORE	+3	Ignoring Parents	-4
Extra Reading	+2	SCREAMING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

Methodology: Why Data-Driven?



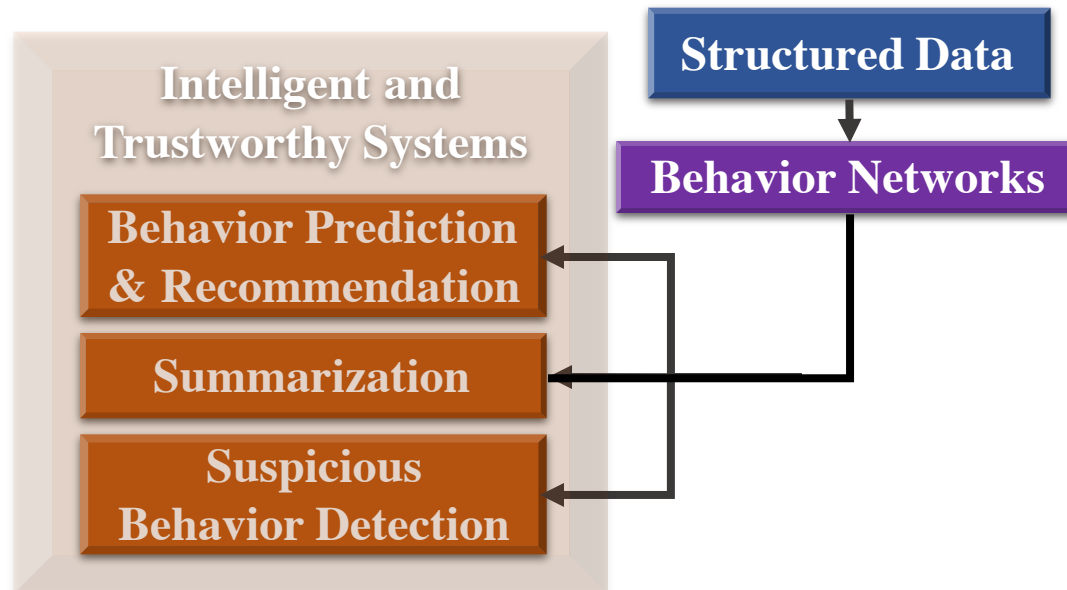
- *Applications.* Recommender systems, fraud/spam detection.
- *Representation.* **Behavior Network** for interaction.
 - **Nodes:** users/authors, items (*e.g.*, products, tweets, papers), *etc.*
 - **Links:** (interaction) following, purchasing, tweeting, publishing, *etc.*
 - **Node attributes:** user profiles, item properties/features, *etc.*
 - **Link attributes:** similarity, distance, weight, *etc.*

Data to Network to Knowledge



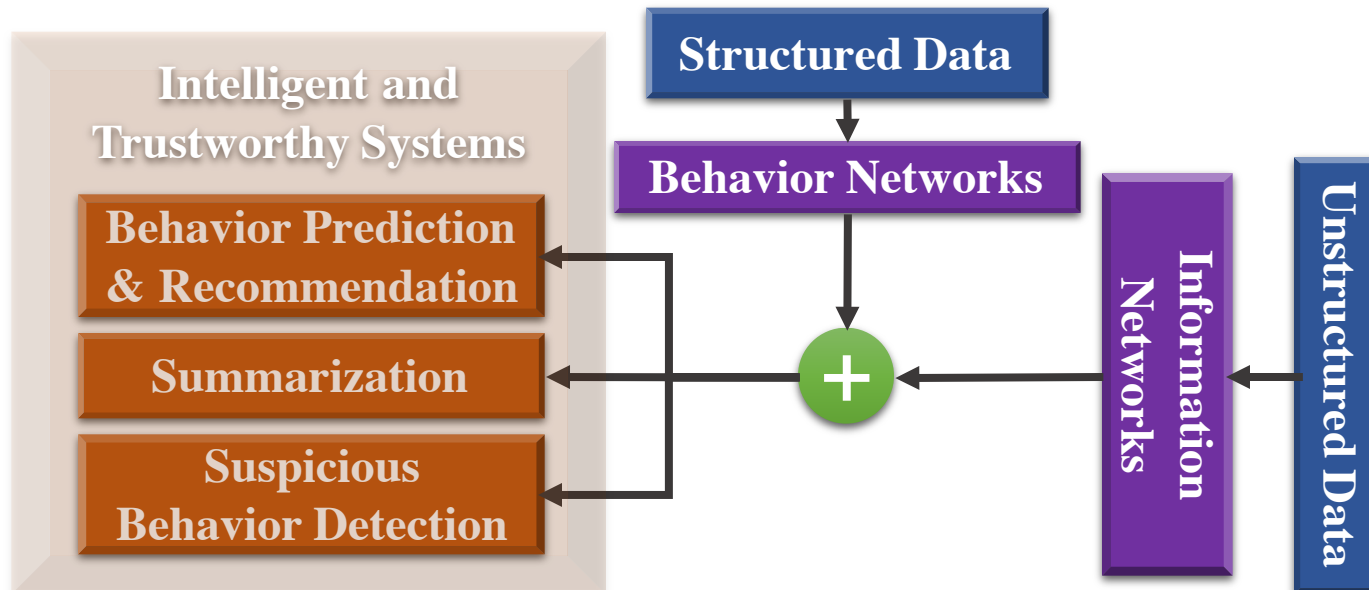
Outline: Data-Driven Behavioral Analytics

- Mining **behavior networks** with **social and spatiotemporal contexts** to support intelligent and trustworthy systems
 - Mining for **behavior prediction and recommendation**
 - Mining for **suspicious behavior detection**



Outline: Data-Driven Behavioral Analytics

- Mining **behavior networks** with **social and spatiotemporal contexts** to support intelligent and trustworthy systems
 - Mining for **behavior prediction and recommendation**
 - Mining for **suspicious behavior detection**
- Structuring behavioral content and integrating behavioral analysis with **information networks**



Acknowledgement



References

- D. Blei, A. Ng, and M. Jordan. “Latent dirichlet allocation.” JMLR, 2003.
- J. Herlocker, J. Konstan, L. Terveen, J. Riedl. “Evaluating collaborative filtering recommender systems.” ACM TOIS, 2004.
- Y. Koren, R. Bell, C. Volinsky. “Matrix factorization techniques for recommender systems.” Computer, 2009.
- Y. Koren. “Factorization meets the neighborhood: A multifaceted collaborative filtering model.” KDD, 2008.
- Y. Koren. “Collaborative filtering with temporal dynamics.” CACM, 2010.
- M. Balabanovic and Y. Shoham. “FAB: Content-based, collaborative recommendation.” CACM, 1997.
- N. Liu and Q. Yang. “Eigenrank: A ranking-oriented approach to collaborative filtering.” SIGIR, 2008.
- N. Liu, M. Zhao, and Q. Yang. “Probabilistic latent preference analysis for collaborative filtering.” CIKM, 2009.

References

- H. Ma, H. Yang, M. Lyu, and I. King. “Sorec: Social recommendation using probabilistic matrix factorization.” CIKM, 2008.
- H. Ma, T. Zhou, M. Lyu, and I. King. “Improving recommender systems by incorporating social contextual information.” ACM TOIS, 2011.
- H. Ma, D. Zhou, C. Liu, M. Lyu, and I. King. “Recommender systems with social regularization.” WSDM, 2011.
- J. Leskovec, A. Singh, and J. Kleinberg. “Patterns of influence in a recommendation network.” PAKDD, 2006.
- P. Massa and A. Paolo. “Trust-aware recommender systems.” RecSys, 2007.
- M. Jamali and E. Martin. “TrustWalker: A random walk model for combining trust-based and item-based recommendation.” KDD, 2009.
- H. Ma, I. King, and M. Lyu. “Learning to recommend with social trust ensemble.” SIGIR, 2009.
- H. Ma, I. King, and M. Lyu. “Learning to recommend with explicit and implicit social relations.” ACM TIST, 2011.

References

- 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.
- B. Hooi, H.A. Song, A. Beutel, N. Shah, K. Shin, and C. Faloutsos. “Fraudar: Bounding graph fraud in the face of camouflage.” KDD, 2016.
- C. Aggarwal and J. Han. “Frequent pattern mining.” Springer, 2014.
- J. Han, J. Pei, B. Mortazavi-Asl, Q. Chen, U. Dayal, and M. Hsu. “FreeSpan: Frequent pattern-projected sequential pattern mining.” KDD, 2000.

References

X. Yan and J. Han. “gspan: Graph-based substructure pattern mining.” ICDM, 2003.

X. Yan and J. Han. “CloseGraph: Mining closed frequent graph patterns.” KDD, 2003.

Y. Sun, J. Han, X. Yan, P.S. Yu, and T. Wu. “PathSim: Meta path-based top-k similarity search in heterogeneous information networks.” VLDB, 2011.

Y. Sun, Y. Yu, and J. Han. “Ranking-based clustering of heterogeneous information networks with star network schema.” KDD, 2009.

Y. Sun, J. Han, P. Zhao, Z. Yin, H. Cheng, and T. Wu. “RankClus: Integrating clustering with ranking for heterogeneous information network analysis.” EDBT, 2009.

Y. Sun, R. Barber, M. Gupta, C. Aggarwar, and J. Han. “Co-author relationship prediction in heterogeneous bibliographic networks.” ASONAM, 2011.

A. El-Kishky, Y. Song, C. Wang, C.R. Voss, and J. Han. “Scalable topical phrase mining from text corpora.” VLDB, 2014.

J. Liu, J. Shang, C. Wang, X. Ren, and J. Han. “Mining quality phrases from massive text corpora.” SIGMOD, 2015.

References

- X. Ren, A. El-Kishky, C. Wang, F. Tao, C.R. Voss, and J. Han. “Effective entity recognition and typing by relation phrase-based clustering.” KDD, 2015.
- X. Ren, W. He, M. Qu, C.R. Voss, H. Ji, and J. Han. “Label noise reduction in entity typing by heterogeneous partial-label embedding.” KDD, 2016.
- C. Wang, M. Danilevsky, N. Desai, Y. Zhang, P. Nguyen, T. Taula, and J. Han. “A phrase mining framework for recursive construction of a topical hierarchy.” KDD, 2013.
- E.E. Papalexakis, C. Faloutsos, N.D. Sidiropoulos. “ParCube: Sparse parallelizable tensor decompositions.” PKDD, 2012.
- D. Koutra, U. Kang, J. Vreeken, and C. Faloutsos. “VOG: Summarizing and understanding large graphs.” SDM, 2014.
- R. Gupta, A. Halevy, X. Wang, S.E. Whang, and F. Wu. “Biperpedia: An ontology for search applications.” VLDB, 2014.
- M. Yahya, S. Whang, R. Gupta, and A. Halevy. “ReNoun: Fact extraction for nominal attributes.” EMNLP, 2014.
- A. Halevy, N. Noy, S. Sarawagi, S.E. Whang, and X. Yu. “Discovering structure in the universe of attribute names.” WWW, 2016.

References

Q. Li, Y. Li, J. Gao, B. Zhao, W. Fan, and J. Han. “Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation.” SIGMOD, 2014.

Q. Li, Y. Li, J. Gao, L. Su, B. Zhao, M. Demirbas, W. Fan, and J. Han. “A confidence-aware approach for truth discovery on long-tail data.” VLDB, 2014.

F. Ma, Y. Li, Q. Li, M. Qiu, J. Gao, S. Zhi, L. Su, B. Zhao, H. Ji, and J. Han. “Faitcrowd: Fine grained truth discovery for crowdsourced data aggregation.” KDD, 2015.

Y. Li, J. Gao, C. Meng, Q. Li, L. Su, B. Zhao, W. Fan, and J. Han. “A survey on truth discovery.” KDD Explorations Newsletter, 2016.

S. Zhi, B. Zhao, W. Tong, J. Gao, D. Yu, H. Ji, and J. Han. “Modeling truth existence in truth discovery.” KDD, 2015.

S. Kumar, R. West, and J. Leskovec. “Disinformation on the Web: Impact, characteristics, and detection of Wikipedia hoaxes.” WWW, 2016.

S. Kumar, F. Spezzano, and V.S. Subrahmanian. “Identifying malicious actors on social media.” ASONAM, 2016. (tutorial)

Thank you!

**Data-Driven Behavioral Analytics:
Observations, Representations and Models**