

# Research Statement: Data-Driven Behavioral Analytics with Networks

Meng Jiang, University of Illinois at Urbana-Champaign  
<http://www.meng-jiang.com>

Behavior is defined as the *interaction* of individuals with themselves and with their environment.<sup>1</sup> Thanks to Information Technologies, online human behaviors are broadly recorded at an unprecedented level. This gives us an opportunity for getting insights into behaviors and our societies from large-scale real data. The Department of Defense (DoD) listed *Computational Modeling of Human Behavior* as one of the six high-priority topics in DoD Basic Research.<sup>2</sup> In order to provide a fundamental understanding and predictive capability of human behavior dynamics from individuals to societies, behavioral analysis has to face the following four complexities: (1) human behaviors are highly dependent on *social contexts*; (2) only by modeling *spatiotemporal contexts*, can we understand when, where and how the behavior happens; (3) behavioral intentions including monetary incentives and other suspicious purposes are a nontrivial part of behavior modeling; and (4) user preferences and sentiments on *unstructured content* are important driving factors. Taking advantage of the massive amount of available behavioral data, I have been developing and promoting novel data-driven approaches that leverage the methodology of *observation*, *representation*, and *modeling* using real data at a scale large enough to make manual analysis and inspection completely impractical in building intelligent and trustworthy systems (e.g., recommender systems and anti-fraud/anti-spam systems). I focus on analyzing behavioral data with network models and algorithms which rely on little or no human annotations for themes including (T1) mining multidimensional behavior networks with social spatiotemporal contexts, (T2) structuring behavioral content into heterogeneous information networks of entities and attributes, and (T3) integrating behavior networks and information networks for accuracy and interpretability.

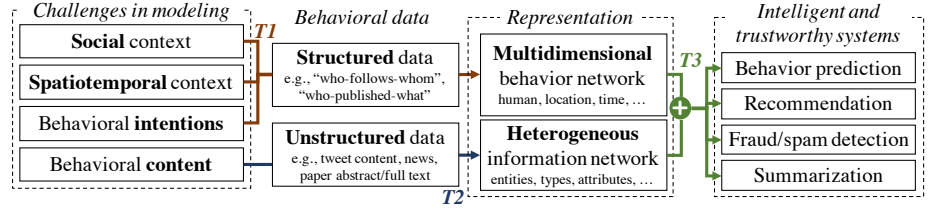


Figure 1: Towards building intelligent and trustworthy systems and addressing the four challenges, I have been proposing data-driven approaches of discovering rich knowledge from behavioral data by *mining* behavior networks (T1), *constructing* information networks (T2) and *integrating* them into in-depth behavioral analysis (T3).

## I. THESIS AND POSTDOCTORAL WORK

My thesis titled “Modeling Complex Behaviors in Social Media” focuses on Theme T1 of understanding behaviors under social spatiotemporal contexts, and my two-year postdoctoral research is broken down into T2 and T3 with contributions mostly to in-depth analysis with unstructured data. The three tasks form the full picture of “Data-Driven Behavioral Analytics with Networks”.

### T1. Mining behavior network with social spatiotemporal contexts

During my five-year Ph.D., I was fortunate to join the long-term collaboration between Tencent Weibo (one of the largest Twitter-style websites in China) and Tsinghua, which put me in the unique position of having access to massive real data generated by millions of users. During this collaboration, I tackled real-world problems such as building recommender systems and anti-fraud systems.

#### T1.1 Modeling social spatiotemporal contexts for behavior prediction

Weibo suffers from *low conversion rate*: Users received too much (less interesting) information from crowds and generated fewer than 6 retweets for every 100 news feed requests. Therefore, can we recommend tweets by predicting their behaviors to address the issue of information overload? Although collaborative filtering techniques have been widely used, we know little about why users adopt or reject items, which has been the bottleneck for further improving the recommendation performance. How do we fully exploit social contextual information from “who-posts-what” (the behavior network) and “who-follows-whom” (the social network) to reveal the information adoption mechanism, and accordingly propose a method for recommendation?

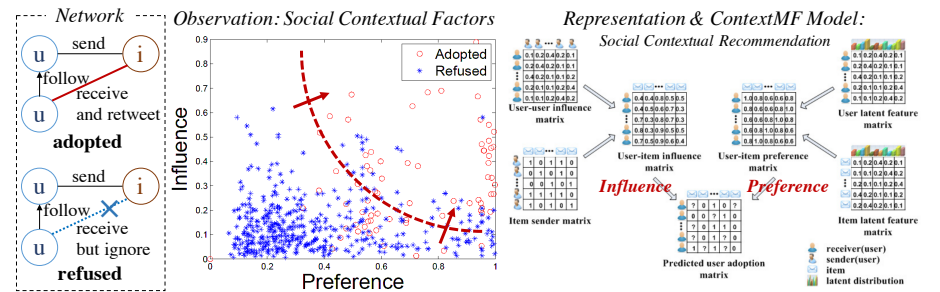


Figure 2: Uncovering two significant factors of user decisions with data, we proposed CONTEXTMF [1] that integrates preference and influence in social recommendation.

To answer this question, we observed from the real network data that personal preference on posted content and interpersonal influence from the content poster are two significant factors that determine users’ decisions (see Figure 2). We proposed a probabilistic matrix factorization model, CONTEXTMF [1], to fuse the behaviors and social contexts. In this model, there are three low-rank latent spaces respectively corresponding to the user space, item space and influence space. They are regularized by the observed preference similarity matrix, content similarity matrix and interaction frequency matrix. The product of the user space and item space corresponds to the preference; the product of influence and preference is proportional to the probability of item adoption. Empirical results on the Weibo dataset demonstrate that CONTEXTMF reduces the root-mean-square error (RMSE) of prediction by 21%.

<sup>1</sup>Behavior – Wikipedia, the free encyclopedia: <https://en.wikipedia.org/wiki/Behavior>

<sup>2</sup>National Defense Industrial Association (NDIA) Conference Proceedings: <http://www.dtic.mil/ndia/>

**Cross-domain behavior modeling.** In principle, if we have denser behavioral data in the form of user-item links we should be able to acquire more knowledge and thus predict more accurately. However, with this method we have not yet addressed the issue of sparsity, where we must “cold start” recommendation for those users who have not yet adopted or rejected many items. Fortunately, Weibo enables users generating content in multiple domains such as profiling tags, images, and interest groups. We developed Hybrid Random Walk (HYBRIDRW) [2], a method which facilitates social recommendation by transferring knowledge across multiple domains on a star-structured network (see Figure 3a). HYBRIDRW collaboratively integrates multiple domains to discover the common knowledge of user tie strength in the central social domain and accordingly alleviates the sparsity issue in individual domains. Empirical results show that our method requires 65% fewer training examples for equivalent baseline performance when it integrates profiling tag information from an auxiliary domain.

**Cross-platform behavior modeling.** In real world, users are active on multiple platforms. So can we utilize rich social platform data to improve the accuracy of predicting behaviors in other platforms (e.g., movie ratings in Netflix, ride requests in Uber)? The natural bridge across platforms is the small number of partially overlapping users, for example, users can register with Uber using their Facebook accounts. Leveraging these bridge users effectively is an open and challenging problem (see Figure 3a). We proposed XPTRANS [3], a method which transfers knowledge across platforms by jointly optimizing different user and item spaces on said platforms and using pairwise similarity of the overlapping users to constrain the spaces. By transferring via the very rare set of users with activity on multiple platforms (accounting for less than 1.1% of all users in our dataset), the predictive performance on the users with sparse activity is even better than the non-transfer predictive performance on highly active users.

**Spatiotemporal behavior modeling.** Human behavior is a product of and evolves with the changing of a multitude of interrelated factors such as physical environment, social identity, and interaction. In a tweet-like setting, we use high-order tensor sequences to represent dynamic multi-faceted behaviors, including (1) user who posts, (2) user who is mentioned, (3) word, and (4) location. We proposed a Flexible Evolutionary Multi-faceted Analysis (FEMA) method [4] based on a dynamic scheme of tensor factorization in which flexible regularizers of social-relationship and location-distance information are imposed to alleviate inherent difficulties from highly sparse tensors (see Figure 3b). To address the issue of complexity, we give approximation algorithms based on Tensor Perturbation theory to factorize the updated tensor with sparse increments, where the loss bound is theoretically proved. Experimental results demonstrate that our method achieves 30.8% higher accuracy when it considers multi-facets and 17.4% higher accuracy when it uses the regularizers. Moreover, it can reduce the time cost from hours to minutes.

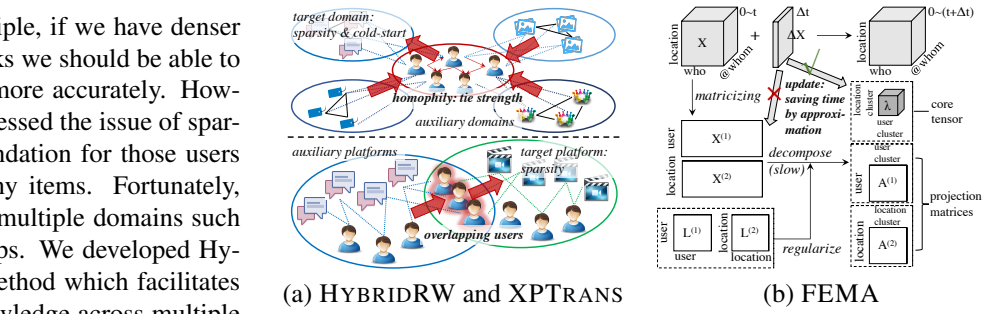
### Impact.

- CONTEXTMF [1] and HYBRIDRW [2] are the 3<sup>rd</sup> and 9<sup>th</sup> most cited papers of CIKM 2012 with **122** and **47** citations among 146 accepted papers at the acceptance rate of 13.4%, where as the median number of citations for CIKM 2012 is 11. Adding the feature of scalability, CONTEXTMF+ [5] has appeared in TKDE 2014 with **36** citations.
- CONTEXTMF+ has been deployed in Weibo’s recommender system. Online testing demonstrates that the conversion rate is improved from 5.78% to 8.27% (relatively **43%**). Note that the rate is strongly related to the company’s ad revenue.

### T1.2 Modeling social spatiotemporal contexts for suspicious behavior detection

Given “who-follows-whom” networks, how can we automatically detect fake followers with high recall? The fraudsters are paid to given certain accounts many additional followers, making them seem more legitimate or famous. Twitter admitted 5% of its users are fake;<sup>3</sup> Weibo suffered a more serious variation of the same problem. Existing approaches learn classifiers with features of potential fraudsters such as the numbers of followees, hashtags, and URLs. However, the fraudsters are smart: They can but they don’t have to follow many or post a lot; some of them even post nothing.

Essentially, we reconsider the detection problem by revealing the fraudsters’ manipulation on the network that they have to do. As shown in Figure 4, the spikes on the out-degree distribution indicate millions of anomalous nodes on the network. The fraudsters exhibit behavior that is (1) synchronized (they often connect to the very same 20, 100 or 500 targets), and (2) abnormal (their behavior pattern is different from the majority of nodes). We developed a scalable and parameter-free algorithm, CATCHSYNC [6], to measure the two properties (synchronicity and normality) of groups of followees. For a follower, the synchronicity score measures how the followees are similar with each other, and the normality score measures how the followees are similar with other nodes in



(a) HYBRIDRW and XPTRANS (b) FEMA  
Figure 3: Contextual behavior modeling: (a) bridging behaviors across domains and platforms, (b) modeling spatiotemporal information with Flexible Evolutionary Multifaceted Analysis.

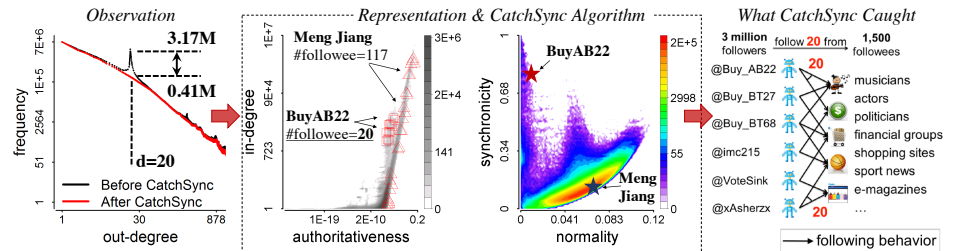


Figure 4: CATCHSYNC [6] spots that fake followers consistently connect to customers of similar features: Their synchronized behaviors create spikes on degree distributions.

<sup>3</sup>Business Insider. <http://www.businessinsider.com/5-of-twitter-monthly-active-users-are-fake-2013-10>

the network. We proved that given the normality score, the synchronicity score has a parabolic lower bound. CATCHSYNC detects millions of fake followers who have unexpectedly high synchronicity and successfully recovers the distribution into a power-law shape, which demonstrates high recall of the performance.

*Evaluating suspiciousness across dimensions.* Besides inflating the number of fans, Weibo’s fraudsters are trying to manipulate the popularity of trending topics. So which of the two seems more suspicious: 5,000 tweets from 200 users on 5 IP addresses, or 10,000 tweets from 500 users on 500 IP addresses but all with the same hashtag and all in 10 minutes? The literature has many methods that find dense blocks in matrices and tensors, but no method gives a principled way to score the suspiciousness of dense blocks with different numbers of dimensions such as user, hashtag, location and time. To address this issue, we formulated axioms that any metric of suspiciousness should satisfy, and proposed a principled and computationally-efficient metric that satisfies the axioms. We further developed CROSSPOT [7] to spot suspicious regions and sort them in importance order. Empirical results show that CROSSPOT successfully detects hashtag-hijacking and retweet-boosting in Weibo datasets spanning 0.3 billion posts, improving performance (measured by F1 score) by 68%.

### Impact.

- CATCHSYNC [6] was selected as one of the best paper finalists of KDD 2014 and currently has **40** citations. It has been taught in (1) CMU 15-826 “Multimedia Databases and Data Mining”, (2) UMich EECS 598 “Graph Mining and Exploration at Scale”, and (3) ASONAM 2016 Tutorial “Identifying Malicious Actors on Social Media” by Kumar et al. from Univ. of Maryland.
- LOCKINFER [8] was the first paper to introduce the concept of *Camouflage* in fraud detection, and it now has **27** citations and it has appeared in KAIS 2015 [9]. CROSSPOT [7] has **16** citations and it has appeared in TKDE 2016 [10].
- Our survey paper about current trends and future directions on suspicious behavior detection has appeared in IEEE Intelligent Systems [11]. It was ranked as the top 10 most frequently downloaded documents in the IS from January to March 2016.

## T2. Structuring behavioral content into heterogeneous information network

In Theme T1, I discussed modeling social spatiotemporal contexts in behavioral data: The contexts are naturally well-structured and thus can be represented as multi-dimensional behavior networks. However, the behavioral content, especially the text corpus in tweets and papers, is information-rich but *unstructured*. Structuring the text into heterogeneous information networks enables deep understanding of the behavioral content. Given the text data, e.g., a news corpus with sentence #1 in Figure 5, can we extract (1) the  $\langle \text{entity type, attribute name} \rangle$  pair, e.g.,  $\langle \text{\$COUNTRY, president} \rangle$  and (2) the  $\langle \text{entity, attribute name, attribute value} \rangle$  tuple, e.g.,  $\langle \text{Burkina Faso, president, Blaise Compaoré} \rangle$ ?

Google mines users’ fact-seeking queries (e.g., “president of united states”) with “*E-A*” patterns (*E* for entity and *A* for attribute name) such as “*A of E*” and “*E’s A*”. However, query-log word distributions are highly constrained compared with ordinary written language. Thus “*E-A*” patterns like “*E A*” and “*A, E*” which lack attribute values will generate noisy attribute name extractions. The state-of-the-art open IE systems learn syntactic and lexical patterns of expressing relationships for extracting entity-level tuples. However, they may generate incorrect or imprecise extractions (e.g.,  $\langle \text{President Blaise Compaoré, have, government of Burkina Faso} \rangle$  in which “have” is not a good attribute name for the entity by ignoring the type information).

Our idea is to investigate joint extraction of the type-level  $\langle \text{entity type, attribute name, attribute value type} \rangle$  tuples (adding “value type” into the pair) and the entity-level tuples because these two extraction processes can mutually enhance each other. First, extracting a frequent type-level tuple is easier (due to its higher frequency) than extracting a concrete entity-level tuple. Second, the type distributions of the concrete entities indicate the appropriate level of the entity types in the type-level tuples. This joint extraction helps find the entity-level tuples with low counts to improve recall if the type-level tuples are frequent and attach attribute names to appropriate levels of types to improve precision. Based on this, we propose a novel methodology, *Meta Pattern Mining*, that mines frequent and informative semantic patterns indicating type-level tuples, called *meta patterns*. For sentence #3 in Figure 5, if we replace “U.S.” with  $\text{\$COUNTRY}$  and replace “Barack Obama” with  $\text{\$POLITICIAN}$ , and carefully segment the sentence, we can generate meta pattern “ $\text{\$COUNTRY president \$POLITICIAN}$ ” and tuple  $\langle \text{\$COUNTRY, president, \$POLITICIAN} \rangle$ . Taking the meta pattern back to the text, we can extract a set of entity-level tuples. Note that mining meta patterns allows *automatic* and *domain-independent* attribute discovery from massive corpora, without making any linguistic assumptions and relying on neither query logs nor human annotations.

Our methodology of mining meta patterns has the advantages described above, but it is challenging. First, imprecise boundaries of meta patterns may generate incorrect or incomplete extractions. Second, adjusting the granularity at which we should type the entities mentioned in a given meta pattern is nontrivial. We develop a *Meta Pattern-driven Attribute Discovery* (METAPAD) framework, which first pre-processes text data to extract entities and their types as input in a data-driven and distantly-supervised manner and

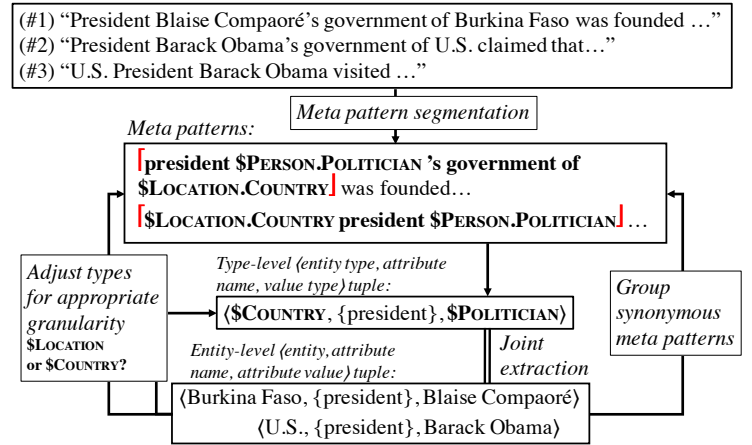


Figure 5: Structuring text into heterogeneous information networks: METAPAD jointly extracts  $\langle \text{entity type, attribute name, value type} \rangle$  tuples and  $\langle \text{entity, attribute name, attribute value} \rangle$  tuples for high precision and high recall. It is automated and domain-independent.

then mines the meta patterns with new techniques to address the challenges for attribute discovery. Experiments on news, tweets and biomedical text data demonstrate that METAPAD improves the F1 scores over the state-of-the-art by 32–51% in ⟨entity type, attribute name⟩ extraction and by 26–35% in ⟨entity, attribute name, attribute value⟩ extraction.

#### Impact.

- In METAPAD [12] I collaborated with Dr. Taylor Cassidy, Dr. Lance M. Kaplan, and Dr. Timothy R. Hanratty from U.S. Army Research Lab (ARL). It is being transferred to the ARL’s Network Science Collaborative Technology Alliance project.

### T3. Integrating behavior network and information network for behavior summarization

Fusing structured and unstructured human behavioral data is imperative for in-depth behavioral analysis. Even bringing the quality phrases that were extracted from the unstructured content into behavior modeling has already been valuable and challenging. For example, given phrases and spatiotemporal contexts of tweets, can we automatically detect and summarize events from the multidimensional data? High-order tensor methods assume that every behavior has exactly one value in every dimension; for example, a tweet has one user, one phrase, and one hashtag. However, in real cases, a behavior may have multiple values in a given dimension. With this observation, we decided to represent all tweets with a “two-level matrix”: (1) the first/second level of columns are dimensions/dimensional values, (2) the first/second level of rows are time slices/behaviors. Thus, an event summary can be defined as a set of interesting dimensions and consecutive time slices, a set of interesting values in each selected dimension (on the columns), and a set of behaviors (tweets) in each selected slice (on the rows), which forms a “Tartan” in the matrix (see Figure 6). We developed a propagation method CATCHTARTAN [13] to capture the Tartans in a principled and scalable way: it determines the meaningfulness of every operation of updating the Tartan with the Minimum Description Length (MDL) principle. Empirical results show that it outperforms the tensor-based approaches, requires no parameters and provides comprehensive summaries of local events in tweets and research trends in academic data.

		User			Location			Hashtag		Phrase			
		$u_1$	$u_2$	$u_3$	$l_1$	$l_2$	$h_1$	$h_2$	$p_1$	$p_2$	$p_3$	$p_4$	
$t_1$	$i_1$	1	0	0	1	0	0	1	0	1	1	0	
	$i_2$	1	0	0	0	0	0	0	1	0	0	0	
$t_2$	$i_3$	0	1	0	1	0	0	1	0	2	0	0	
	...												
$t_3$	$i_n$	0	0	1	1	0	0	2	0	1	1	0	
	...												

Florida #PrayingForHaiti “please\_donate” “those in need”

Figure 6: Each row is a tweet. [13] detects and summarizes events by looking for *Tartans* in a “two-level matrix” which represents every dimension of tweets including time, location and *phrase*.

#### Fun fact.

- CATCHTARTAN [13] was accepted with oral presentation in KDD 2016 (acceptance rate: 8.9%). This is the 9<sup>th</sup> paper I have collaborated on with Dr. Faloutsos as the 1<sup>st</sup> author after my 9-month visit to CMU.

## II. FUTURE RESEARCH DIRECTIONS

Although areas such as economics, politics, and business intelligence are already incorporating data science, I anticipate that data-driven behavioral analytics will also be incorporated with other fields of science such as psychology and sociology. The need for intelligence, trustworthiness, and scalability in behavior modeling will only increase.

### Mid-Term Research Plan

With respect to my mid-term plan (first 3-5 years), I have provided below a more detailed outline of the directions and challenges that I am planning to tackle, both contributing to data-driven behavioral analysis that supports decision-making processes.

#### D1. Intelligence: Integrating structured and unstructured data for in-depth behavioral analysis

1. *Integrating information networks from behavioral content for contextual behavior modeling.* In my thesis and postdoctoral research, I have improved the state-of-the-art in modeling user behavior with structured data: predicting behaviors with contexts and catching suspicious behaviors. However, if we don’t model the content with structures, big data can turn into a big mess. Fortunately, the data-driven meta-pattern mining approaches facilitate automatically structuring the content into a rich network of entities and attributes. I still need to investigate ways to leverage this opportunity to develop accurate and interpretable predictive models.

2. *Predicting behavioral contexts over predicting contextual behaviors.* Suppose we are able to predict with 100% accuracy whether a behavior (e.g., paper-publishing) will occur given its full contexts (e.g., co-authors, datasets, problems, methods). Then, given a subset of the contexts (e.g., the datasets and problems), can we predict the rest of its contexts? For example, if one wants to solve the fraud detection problem on Weibo, which experts, papers, and algorithms should one find, read, and try? Supposing that we have represented the application problems, datasets and other entities as nodes in an  $n$ -node heterogeneous network, the complexity of searching for the optimum is too high: the space of solutions is  $O(n!)$ . Reducing the complexity into a practical level is a challenging problem: One possible solution is to prune impossible permutations of the contexts with statistics and insights from social, behavioral, and psychological scientific studies.

#### D2. Trustworthiness: Structuring reliable information networks from behavioral content

Our meta-pattern mining is a general, data-driven methodology with applications to be used across multiple NLP tasks such as entity recognition and typing, attribute/relationship extraction, and slot filling in an automated way. However, a portion of the extractions may still be incorrect or imprecise, for example, the text “McDonald’s U.S. President...” may mistake a McDonald executive as a U.S. president. In order to structure reliable information from behavioral content, it is necessary to enhance mining results with majority voting-based conflict resolution, sentence structure-based entity refinement, and conditional functional dependency rule mining.

*Collaborative research.* I am eager for collaborations with experts in Natural Language Processing, Text Mining, Machine Learning, Artificial Intelligence, and Cyber-Physical Security and Systems, working towards the ultimate goal of building efficient



systems of intelligence and trustworthiness for human beings and our society.

## Long-Term Vision: Interdisciplinary Research and Real-World Impact of Behavioral Analytics

The future of behavioral analytics lies in the intersection of data science and behavioral psychology, which is a formidable combination wherein each group brings to the table unique skills that differ by scientific training.

1. *Bringing psychological expertise into data technology.* Data in itself is nothing. We can only create value if we turn it into information. Data expertise alone is not sufficient to truly understand human behaviors. Behavioral scientists, who typically specialize in cognitive psychology, know more about the human brain. How do people perceive? How do they reach a decision? This kind of knowledge is crucial if we want to interpret users' behavior. Take the social contextual recommendation as an example. When I was struggling to understand how Weibo users decide whether to forward or ignore a message, a Science paper [14] emerged to help. Salganick et al. adopted an experimental approach to the study of social influence in cultural markets, an approach which inspired me to reconsider the information adoption mechanism of Weibo, leading me to eventually propose that preference and influence were two major contextual factors. Experiments on real large data validated my assumption (see Figure 2) and the proposed model CONTEXTMF significantly improved Weibo's recommender systems. It is important to realize that vision if we bring real psychological expertise into our data science.

2. *Facilitating psychological discovery processes with data science.* Data-driven approaches have enormous potential to change the way psychologists observe human behavior. Big data leads researchers to a point where they can collect behavioral information without sampling human participants at all. Social media sites such as Facebook and Twitter are the new "macroscopes of human behavior." Technology such as smart phones and wearable sensors can gather information on physical activity, social interaction, and so on. This offers us a clue to understand psychological principles and behavioral theories for experimentation. Psychologists and data scientists have a vested interest in working together to yield more explanatory insight that can change the world.

3. *Deploying scalable behavior modeling for real systems.* To scale complex behavior models to the high volume of data, my research efficiently exploits the structure of multidimensional behavior networks and heterogeneous information networks. The direct impact of my work on Tencent Weibo's revenue stream is compelling evidence for the value of evolutionary analysis (e.g., [5, 4]) and scalable algorithms (e.g., [6, 13]) across a wide variety of applications.

*Contributing to the research community.* I am focused on studying and promoting Data-Driven Behavioral Analytics. I have co-authored two 3-hour tutorials in major conferences (i.e., ICDM 2015 [15] and CIKM 2016 [16]). Both tutorials were well attended, and I received \$700 honorarium from the ICDM 2015. I also have written two book chapters about user behavior modeling [17, 18].

*Funding opportunities.* First, based on the idea of CATCHTARTAN [13], I contributed significantly to the technical content in the proposal "NSF III: Small: Multi-Dimensional Structuring, Summarizing and Mining of Social Media Data." It has been awarded to the PI Dr. Jiawei Han by NSF IIS (08-01-2016 to 07-31-2019, \$500,000). I am the major supported member.<sup>4</sup> Besides "Information Integration and Informatics" (III), "Secure and Trustworthy Cyberspace" (SaTC) also encourages the study of behavior analysis. Second, I have extensively collaborated with U.S. Army Research Lab (ALC and APG) in the NS CTA projects. Third, I have international relations with China's institutes (e.g., Tsinghua, Microsoft Research Asia) and IT companies (e.g., Tencent, Alibaba).

## REFERENCES

- [1] **Meng Jiang**, Peng Cui, Rui Liu, Qiang Yang, Fei Wang, Wenwu Zhu, and Shiqiang Yang. Social contextual recommendation. In *ACM CIKM*, 2012.
- [2] **Meng Jiang**, Peng Cui, Fei Wang, Qiang Yang, Wenwu Zhu, and Shiqiang Yang. Social recommendation across multiple relational domains. In *ACM CIKM*, 2012.
- [3] **Meng Jiang**, Peng Cui, Nicholas Jing Yuan, Xing Xie, and Shiqiang Yang. Little is much: Bridging cross-platform behaviors through overlapped crowds. In *AAAI*, 2016.
- [4] **Meng Jiang**, Peng Cui, Fei Wang, Xinran Xu, Wenwu Zhu, and Shiqiang Yang. Fema: Flexible evolutionary multi-faceted analysis for dynamic behavioral pattern discovery. In *ACM SIGKDD*, 2014.
- [5] **Meng Jiang**, Peng Cui, Fei Wang, Wenwu Zhu, and Shiqiang Yang. Scalable recommendation with social contextual information. *IEEE TKDE*, 2014.
- [6] **Meng Jiang**, Peng Cui, Alex Beutel, Christos Faloutsos, and Shiqiang Yang. Catchsync: Catching synchronized behavior in large directed graphs. In *ACM SIGKDD (best paper finalist)*, 2014.
- [7] **Meng Jiang**, Alex Beutel, Peng Cui, Bryan Hooi, Shiqiang Yang, and Christos Faloutsos. A general suspiciousness metric for dense blocks in multimodal data. In *IEEE ICDM*, 2015.
- [8] **Meng Jiang**, Peng Cui, Alex Beutel, Christos Faloutsos, and Shiqiang Yang. Inferring strange behavior from connectivity pattern in social networks. In *PAKDD*, 2014.
- [9] **Meng Jiang**, Peng Cui, Alex Beutel, Christos Faloutsos, and Shiqiang Yang. Inferring lockstep behavior from connectivity pattern in large graphs. *KAIS*, 2015.
- [10] **Meng Jiang**, Alex Beutel, Peng Cui, Bryan Hooi, Shiqiang Yang, and Christos Faloutsos. Spotting suspicious behaviors in multimodal data: A general metric and algorithms. *IEEE TKDE*, 2016.
- [11] **Meng Jiang**, Peng Cui, and Christos Faloutsos. Suspicious behavior detection: Current trends and future directions. *IEEE Intelligent Systems*, 2016.
- [12] **Meng Jiang**, Jingbo Shang, Taylor Cassidy, Xiang Ren, Lance Kaplan, Timothy Hanratty, and Jiawei Han. Metapad: Meta pattern-driven attribute discovery in massive text corpora. In *Anonymized (in peer review)*, 2017.
- [13] **Meng Jiang**, Christos Faloutsos, and Jiawei Han. Catchtartan: Representing and summarizing dynamic multicontextual behaviors. In *ACM SIGKDD*, 2016.
- [14] Matthew J Salganik, Peter Sheridan Dodds, and Duncan J Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *science*, 311(5762):854–856, 2006.
- [15] **Meng Jiang** and Peng Cui. Behavior modeling in social networks: From micro to macro. *IEEE ICDM (tutorial)*, 2015.
- [16] **Meng Jiang**, Peng Cui, and Jiawei Han. Data-driven behavioral analytics: Observations, representations and models. *ACM CIKM (tutorial)*, 2016.
- [17] **Meng Jiang** and Peng Cui. Mining user behaviors in large social networks. *Big Data in Complex and Social Networks*, 2016.
- [18] **Meng Jiang**. Behavior modeling in social networks. *Encyclopedia of Social Network Analysis and Mining*, 2nd edition, 2016.

<sup>4</sup>NSF IIS 16-18481. [http://hanj.cs.illinois.edu/projs/social\\_media.htm](http://hanj.cs.illinois.edu/projs/social_media.htm)