KDD 2017 Tutorial: Data-Driven Approaches towards Malicious Behavior Modeling

Meng Jiang[1], Srijan Kumar[2], V.S. Subrahmanian[2], Christos Faloutsos[3]
[1]Computer Science Department, University of Illinois Urbana-Champaign, IL, USA
[2]Computer Science Department, University of Maryland, College Park, MD, USA
[3]Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA
mjiang89@illinois.edu, srijan@cs.umd.edu, vs@umiacs.umd.edu, christos@cs.cmu.edu

ABSTRACT

The safety, reliability and usability of web platforms are often compromised by malicious entities, such as vandals on Wikipedia, bot connections on Twitter, fake likes on Facebook, and several more. Computational models developed with large-scale real-world behavioral data have shown significant progress in identifying these malicious entities. This tutorial discusses three broad directions of state-of-the-art data-driven methods to model malicious behavior: (i) *feature-based algorithms*, in which distinguishing behavioral features are proposed to predict the malicious users; (ii) *spectral-based algorithms*, which have been widely used in settings of directed graphs, undirected graphs, and bipartite graphs such as "who-follows-whom" Twitter data and "who-likes-what" Facebook data; and (iii) *density-based algorithms*, which efficiently look for suspicious, highly-dense components in multi-dimensional behavioral data. This tutorial will introduce the details of the general algorithms from the above three classes that can be applied to any platform and dataset.

ACM Reference format:
Meng Jiang, Srijan Kumar, V.S. Subrahmanian, and Christos Faloutsos. 2017. KDD 2017 Tutorial: Data-Driven Approaches towards Malicious Behavior Modeling. In *Proceedings of ACM SIGKDD conference, Halifax, Nova Scotia, Canada, August 2017 (SIGKDD'17),* 4 pages.
DOI: 10.475/123_4

1 INTRODUCTION

The web provides a unifying platform for people around the world. This has led to several ground-breaking advancements and positively impacted lives of billions of people. At the same time, it also gives an avenue for entities to engage in malicious behavior [36]. Such malicious users and their activities compromise the integrity of web platforms and make the web experience unsafe and unusable. For instance, it is estimated that a considerable fraction of online entities are malicious — 8%–10% social network accounts and 16% Yelp reviews are fake [1, 26, 35], and 3%–4% of Wikipedia editors are vandals [30]. The implications of malicious entities is far-reaching, as 73% of internet users have witnessed online harassment, and 40% have experienced it themselves [2]. Therefore, it is imperative to identify these malicious users efficiently and as soon as possible.

In this tutorial, we discuss computational algorithms to predict malicious users in large-scale web platforms. These algorithms can be categorized into three complementary directions — feature-based, spectral-based and density-based.

Feature-based algorithms model the users and activities by representing them using a set of attributes, such that each entity is represented as a point in a multi-dimentional space. In order to distinguish benign from malicious users, appropriate set of attributes should be engineered, so that different entities lie in significantly different regions in this feature space. This class of algorithms often requires pre-labeled training data, which are used to find the distinguishing appropriate attributes. These algorithms have been successfully used to predict trolls [8], vandals [30], hoaxes [32], bots [43], sockpuppets [28], among several other malicious entities.

Spectral-based algorithms project user behavioral data into spectral subspaces usually by singular vector decomposition (SVD), when the data can be represented as adjacency matrices of directed graphs, undirected graphs, and bipartite graphs. The spectral subspace plots show suspicious patterns if the graph has malicious entities that exhibit one or more strange lockstep behaviors (e.g., "blocks", "staircases", and "camouflage") [16, 22, 39]. These algorithms do not necessarily require training labels and are computationally fast.

Density-based algorithms are designed based on the intuition that malicious entities act synchronously, i.e. they often take similar action in near-similar time. Benign entities, on the other hand, are not as synchronous. When representing entities using their actions in a multi-dimentional space, malicious entities form a 'dense-block', as an artifact of the synchronous behavior. Density-based algorithms are aimed at identifying these dense-blocks in large-scale behavior logs. These algorithms have successfully been used to predict ill-gotten page Likes [4], zombie followers [21], and social spam [19].

2 TUTORIAL OUTLINE

We will discuss the three categories of algorithms sequentially, starting with feature-based algorithms, then moving to spectral-based algorithms, and finally concluding with density-based algorithms. A brief outline of the tutorial is given below:

Introduction: (15 minutes)

- What is malicious behavior? Why do we model malicious behaviors [44]?
- What are the current trends in malicious behavior modeling [23]?

SIGKDD'17, Halifax, Nova Scotia, Canada

DOI: 10.475/123_4

- A brief introduction to the three major directions of data-driven algorithms — feature-based, spectral-based and density-based algorithms.

Feature-based algorithms: (75 minutes)
- Statistical analysis for feature development.
- Detection of antisocial users: bots [43], trolls [7, 29], sockpuppets [5, 28] and vandals [14, 30].
- Detection of misinformation: hoaxes [32] and rumors [12, 17, 40].

Break (30 minutes)

Spectral-based algorithms: (40 minutes)
- Introduction to "spokes" in spectral subspace plots as indicators of strange behaviors [39].
- Detection of "blocks", "staircases", and "camouflage" with spectral subspace plots [22, 41].
- Theoretical guarantees for preventing "camouflage" [16].

Density-based algorithms: (40 minutes)
- Detection of temporal bipartite cores [4].
- Detection of synchronized behaviors [13, 21].
- Evaluation of suspiciousness of behaviors across multiple dimensions [19, 42].
- Summarization of dynamic and multicontextual behaviors [25].

Conclusion and future directions: (10 minutes)

Additionally, we will also refer to the following works in this tutorial: (1) fake account detection [6, 9–11, 33, 34, 46], (2) fake review detection [15, 37, 38, 45], and (3) social spam detection [18].

3 TARGET AUDIENCE AND PREREQUISITE

This tutorial targets academic, industry and government researchers and practitioners with interests in user behavior modeling, social network anomaly detection, graph mining, and cybersecurity. Beginners in the area will learn the basics of these data-driven approaches. Experts in the area will learn in-depth algorithms and models to detect malicious behaviors. This tutorial should appeal to researchers of several disciplines, and should attract strong attendance at ACM SIGKDD 2017.

There are no prerequisites for attending the tutorial. We will cover basics as well as advanced techniques.

4 TUTORS' INFORMATION

Meng Jiang (corresponding author)
Affiliation: University of Illinois at Urbana-Champaign
E-mail: mjiang89@illinois.edu
Address: Rm 2130, Siebel Center for Computer Science, 201 N. Goodwin Avenue, Urbana, IL 61801, USA
Phone: 217-418-6072

Srijan Kumar
Affiliation: University of Maryland, College Park
E-mail: srijan@cs.umd.edu
Address: A.V. Williams Building, University of Maryland, 8223 Paint Branch Drive, College Park, MD 20742, USA
Phone: 301-329-1103

V.S. Subrahmanian
Affiliation: University of Maryland, College Park
E-mail: vs@umiacs.umd.edu
Address: A.V. Williams Building, University of Maryland, 8223 Paint Branch Drive, College Park, MD 20742, USA
Phone: 301-405-6724

Christos Faloutsos
Affiliation: Carnegie Mellon University
E-mail: christos@cs.cmu.edu
Address: GHC 8019, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213-3891, USA
Phone: 412-268-3505

5 TUTORS' BIO AND EXPERTISE

Meng Jiang a Postdoctoral Research Associate in the Computer Science Department at University of Illinois at Urbana-Champaign. His research interests focus on data-driven behavioral analytics for prediction, recommendation, and suspicious behavior detection. He obtained his Ph.D. and Dissertation Award in 2015 and B.E. in 2010 from Tsinghua University, China. He visited Carnegie Mellon University in 2013, and he visited University of Maryland, College Park in 2016. He has published over 20 refereed articles and 2 book chapters. He gave tutorials in ICDM'15 and CIKM'16, and both tutorials had strong attendance (50+ and 70+) and won the honorariums. He received the SIGKDD 2014 Best Paper Finalist as the first author. More details can be found at http://www.meng-jiang.com/.

Srijan Kumar Srijan Kumar is a postdoctoral researcher in Computer Science department at Stanford University, working on applied machine learning and social network analysis. He has received several awards including UMD University and Dean fellowships, WorldQuant PhD scholarship, IIT Kharagpur gold medal, and WWW 2017 best paper award honorable mention. He received his Ph.D. in computer science from University of Maryland, College Park and his bachelor's degree in computer science from IIT Kharagpur. More details can be found at http://cs.stanford.edu/~srijan/.

V.S. Subrahmanian is a Professor in the Computer Science department, director of the Lab for Computational Cultural Dynamics and Director of the Center for Digital International Government at the University of Maryland, College Park. His work stands squarely at the intersection of big data analytics for increased security, policy, and business needs. He has published over 280 peer-reviewed papers including papers on detecting bots on Twitter, detecting trolls on Slashdot, and detecting vandals on Wikipedia. He led the team that won DARPA's Twitter Bot Challenge in early 2015. He currently serves on the boards of numerous journals including Science, ACM Transactions on Intelligent Systems & Technology, ACM Transactions on Computational Logic, and IEEE Transactions on Computational Social Systems. Moreover, he serves currently on the Research Advisory Board of Tata Consultancy Services, the Board of Directors of the Development Gateway, Sentimetrix, Inc., and CosmosId. More details can be found at http://cs.umd.edu/~vs/.

Christos Faloutsos is a Professor at Carnegie Mellon University. He has received the Presidential Young Investigator Award by the National Science Foundation (1989), the Research Contributions

Award in ICDM 2006, the SIGKDD Innovations Award (2010), 24 "best paper" awards (including 5 "test of time" awards), and four teaching awards. Six of his advisees have attracted KDD or SCS dissertation awards. He is an ACM Fellow. He has served as a member of the executive committee of SIGKDD; he has published over 350 refereed articles, 17 book chapters and two monographs. He holds seven patents (and 2 pending), and he has given over 40 tutorials and over 20 invited distinguished lectures. His research interests include large-scale data mining with emphasis on graphs and time sequences; anomaly detection, tensors, and fractals. More details can be found at http://www.cs.cmu.edu/~christos/.

6 RELATED PAST TUTORIALS

Our tutors have co-presented multiple tutorials as follows:

(a) Alex Beutel, Leman Akoglu, and **Christos Faloutsos**. "Graph-based user behavior modeling: from prediction to fraud detection", ACM SIGKDD 2015 [3].
This tutorial only focussed on graph-based algorithms, and did not discuss the other two categories of algorithms — feature-based and spectral-based. Additionally, our proposed tutorial covers wider types of malicious users and activities, and their detection.

(b) **Meng Jiang** and Peng Cui. "Behavioral modeling in social networks: from micro to macro", IEEE ICDM 2015 [20].
This tutorial focused on how to leverage the social network information for social recommendation, social relationship prediction, and social spam detection. Our proposed tutorial will focus on modeling the malicious behaviors, which is not limited to the social network applications and social datasets.

(c) **Srijan Kumar**, Francesca Spezzano and **V.S. Subrahmanian**. "Identifying malicious actors on social media", ASONAM 2016 [31].
This tutorial was mostly focused on feature-based algorithms, and did not cover spectral-based and density-based algorithms in detail. It covered detection of bots, trolls, hoaxes and vandals. In addition to these, we are proposing a wider spectrum of malicious behavior in this current tutorial.

(d) **Meng Jiang**, Peng Cui, and Jiawei Han. "Data-driven behavioral analytics: observations, representations and models", CIKM 2016 [24].
This tutorial focused on data-driven approaches for modeling behavioral contexts and behavioral content. It introduced the roles of social and spatiotemporal information (as contexts) in predicting human behaviors. It also covered the state-of-the-art text mining techniques that mine structures from rich behavioral content. Our proposed tutorial will focus in depth on modeling one type of behavioral intentions – malicious behaviors. We will introduce general methodologies and data-driven approaches towards this specific research problem.

(e) **Srijan Kumar**, Justin Cheng, and Jure Leskovec. "Antisocial behavior on the Web: characterization and detection", WWW 2017 [27].
This tutorial focused on different types of malicious users, rather than on the different categories of algorithms. The papers discussed in this tutorial were specific to the malicious user being discussed and the algorithms were primarily feature-based. The proposed tutorial additionally covers spectral-based and density-based algorithms and is not focused on specific malicious behavior.

This is the first time the highlighted tutors are delivering a tutorial together: we collect algorithms, models, and methods for different applications from the above tutorials, and organize the rich state-of-the-art from the *perspective of methodologies* towards malicious behavior modeling. We will provide not only fundamental studies but also new insights on this direction.

All the above tutorials attracted strong attendance, and therefore, we expect a strong attendance for our proposed tutorial in KDD 2017 as well.

7 EQUIPMENT

The tutorial does not require any special equipment. We will use our own laptop for presentation.

8 SLIDES DUE AND PREVIOUS WEBSITES

We understand the due of slides is on July 31st, 2017. Actually if the proposal is accepted, we will get the slides and website ready before July. Please kindly check our previous websites below.
(b) ICDM'15: http://www.meng-jiang.com/tutorial-icdm15.html
(c) ASONAM'16: http://www.cs.umd.edu/~srijan/badactorstutorial/
(d) CIKM'16: http://www.meng-jiang.com/tutorial-cikm16.html

9 VIDEO SNIPPET

Here are video links of our tutors' talks from videolectures or YouTube. We give at most 3 links for each tutor.

Meng Jiang
KDD'14 CatchSync: http://videolectures.net/kdd2014_jiang_catchsync/
KDD'14 FEMA: http://videolectures.net/kdd2014_jiang_fema/
KDD'16 CatchTartan: http://videolectures.net/kdd2016_jiang_multicontextual_behaviors/

V.S. Subrahmanian
Cyber Security and Resilience Conference: https://www.youtube.com/watch?v=IAsh9-nrypY
Forecasting Malware Spread in Networks: https://www.youtube.com/watch?v=H_5kPDmwCTo
Winning the DARPA Twitter Bot Challenge: https://www.youtube.com/watch?v=N80EhaIS82k

Christos Faloutsos
Mining large graphs: https://www.youtube.com/watch?v=g7n4_C79oaE
How to find patterns in large graphs: https://www.youtube.com/watch?v=GBzoNgqF-gQ
Mining large graphs, patterns, anomalies, and fraud detection: https://www.youtube.com/watch?v=UyjhxEKjceA

REFERENCES

[1] 2014. How Many Of The Internetfis Users Are Fake. http://www.dailyinfographic.com/how-many-of-the-internets-users-are-fake. (2014).

[2] 2014. Online Harassment, Pew Research Center. http://www.pewinternet.org/2014/10/22/online-harassment. (2014).

[3] Alex Beutel, Leman Akoglu, and Christos Faloutsos. 2015. Graph-Based User Behavior Modeling: From Prediction to Fraud Detection. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2309–2310.

[4] Alex Beutel, Wanhong Xu, Venkatesan Guruswami, Christopher Palow, and Christos Faloutsos. 2013. Copycatch: stopping group attacks by spotting lockstep behavior in social networks. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, 119–130.

[5] Zhan Bu, Zhengyou Xia, and Jiandong Wang. 2013. A sock puppet detection algorithm on virtual spaces. *Knowledge-Based Systems* 37 (2013), 366–377.

[6] Qiang Cao, Michael Sirivianos, Xiaowei Yang, and Tiago Pregueiro. 2012. Aiding the detection of fake accounts in large scale social online services. In *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*. USENIX Association, 15–15.

[7] Justin Cheng, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2017. Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions. (2017).

[8] Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2015. Antisocial Behavior in Online Discussion Communities. In *Proceedings of the Ninth International AAAI Conference on Web and Social Media*.

[9] John P Dickerson, Vadim Kagan, and VS Subrahmanian. 2014. Using sentiment to detect bots on Twitter: Are humans more opinionated than bots?. In *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*. IEEE, 620–627.

[10] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The rise of social bots. *Commun. ACM* 59, 7 (2016), 96–104.

[11] Michael Fire, Gilad Katz, and Yuval Elovici. 2012. Strangers intrusion detection-detecting spammers and fake profiles in social networks based on topology anomalies. *Human Journal* 1, 1 (2012), 26–39.

[12] Adrien Friggeri, Lada A Adamic, Dean Eckles, and Justin Cheng. 2014. Rumor Cascades.. In *ICWSM*.

[13] Maria Giatsoglou, Despoina Chatzakou, Neil Shah, Alex Beutel, Christos Faloutsos, and Athena Vakali. 2015. ND-Sync: Detecting Synchronized Fraud Activities. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 201–214.

[14] Stefan Heindorf, Martin Potthast, Hannah Bast, Björn Buchhold, and Elmar Haussmann. 2017. WSDM Cup 2017: Vandalism Detection and Triple Scoring. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM, 827–828.

[15] Bryan Hooi, Neil Shah, Alex Beutel, Stephan Günnemann, Leman Akoglu, Mohit Kumar, Disha Makhija, and Christos Faloutsos. 2016. Birdnest: Bayesian inference for ratings-fraud detection. In *Proceedings of the 2016 SIAM International Conference on Data Mining*. SIAM, 495–503.

[16] Bryan Hooi, Hyun Ah Song, Alex Beutel, Neil Shah, Kijung Shin, and Christos Faloutsos. 2016. Fraudar: Bounding graph fraud in the face of camouflage. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 895–904.

[17] Benjamin D. Horne and Sibel Adali. 2017. This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News. In *The Second International Workshop on News and Public Opinion*.

[18] Xia Hu, Jiliang Tang, Huiji Gao, and Huan Liu. 2014. Social spammer detection with sentiment information. In *Data Mining (ICDM), 2014 IEEE International Conference on*. IEEE, 180–189.

[19] Meng Jiang, Alex Beutel, Peng Cui, Bryan Hooi, Shiqiang Yang, and Christos Faloutsos. 2015. A general suspiciousness metric for dense blocks in multimodal data. In *Data Mining (ICDM), 2015 IEEE International Conference on*. IEEE, 781–786.

[20] Meng Jiang and Peng Cui. 2015. Behavioral modeling in social networks: from micro to macro. In *International Conference on Data Mining*. IEEE, 2309–2310.

[21] Meng Jiang, Peng Cui, Alex Beutel, Christos Faloutsos, and Shiqiang Yang. 2014. Catchsync: catching synchronized behavior in large directed graphs. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 941–950.

[22] Meng Jiang, Peng Cui, Alex Beutel, Christos Faloutsos, and Shiqiang Yang. 2014. Inferring strange behavior from connectivity pattern in social networks. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 126–138.

[23] Meng Jiang, Peng Cui, and Christos Faloutsos. 2016. Suspicious behavior detection: Current trends and future directions. *IEEE Intelligent Systems* 31, 1 (2016), 31–39.

[24] Meng Jiang, Peng Cui, and Jiawei Han. 2016. Data-driven behavioral analytics: Observations, representations and models. *ACM CIKM (tutorial)* (2016).

[25] Meng Jiang, Christos Faloutsos, and Jiawei Han. 2016. Catchtartan: Representing and summarizing dynamic multicontextual behaviors. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 945–954.

[26] Katharina Krombholz, Dieter Merkl, and Edgar Weippl. 2012. Fake identities in social media: A case study on the sustainability of the Facebook business model. *Journal of Service Science Research* 4, 2 (2012), 175–212.

[27] Srijan Kumar, Justin Cheng, and Jure Leskovec. 2017. Antisocial Behavior on the Web: Characterization and Detection. In *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 947–950.

[28] Srijan Kumar, Justin Cheng, Jure Leskovec, and VS Subrahmanian. 2017. An Army of Me: Sockpuppets in Online Discussion Communities. In *Proceedings of the 26th international conference on World Wide Web*. ACM.

[29] Srijan Kumar, Francesca Spezzano, and VS Subrahmanian. 2014. Accurately detecting trolls in slashdot zoo via decluttering. In *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*. IEEE, 188–195.

[30] Srijan Kumar, Francesca Spezzano, and VS Subrahmanian. 2015. VEWS: A wikipedia vandal early warning system. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 607–616.

[31] Srijan Kumar, Francesca Spezzano, and VS Subrahmanian. 2016. Tutorial: Identifying Malicious Actors on Social Media. In *Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on*. IEEE.

[32] Srijan Kumar, Robert West, and Jure Leskovec. 2016. Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 591–602.

[33] Kyumin Lee, James Caverlee, and Steve Webb. 2010. Uncovering social spammers: social honeypots+ machine learning. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. ACM, 435–442.

[34] Yixuan Li, Oscar Martinez, Xing Chen, Yi Li, and John E Hopcroft. 2016. In a world that counts: Clustering and detecting fake social engagement at scale. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 111–120.

[35] Michael Luca and Georgios Zervas. 2016. Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science* 62, 12 (2016), 3412–3427.

[36] Anne P Mintz. 2002. *Web of deception: Misinformation on the Internet*. Information Today, Inc.

[37] Arjun Mukherjee, Bing Liu, and Natalie Glance. 2012. Spotting fake reviewer groups in consumer reviews. In *Proceedings of the 21st international conference on World Wide Web*. ACM, 191–200.

[38] Arjun Mukherjee, Vivek Venkataraman, Bing Liu, and Natalie S Glance. 2013. What yelp fake review filter might be doing?. In *ICWSM*.

[39] B Aditya Prakash, Ashwin Sridharan, Mukund Seshadri, Sridhar Machiraju, and Christos Faloutsos. 2010. Eigenspokes: Surprising patterns and scalable community chipping in large graphs. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 435–448.

[40] Vahed Qazvinian, Emily Rosengren, Dragomir R Radev, and Qiaozhu Mei. 2011. Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 1589–1599.

[41] Neil Shah, Alex Beutel, Brian Gallagher, and Christos Faloutsos. 2014. Spotting suspicious link behavior with fbox: An adversarial perspective. In *Data Mining (ICDM), 2014 IEEE International Conference on*. IEEE, 959–964.

[42] Kijung Shin, Bryan Hooi, Jisu Kim, and Christos Faloutsos. 2017. D-Cube: Dense-Block Detection in Terabyte-Scale Tensors. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM, 681–689.

[43] VS Subrahmanian, Amos Azaria, Skylar Durst, Vadim Kagan, Aram Galstyan, Kristina Lerman, Linhong Zhu, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. 2016. The DARPA Twitter bot challenge. *Computer* 49, 6 (2016), 38–46.

[44] VS Subrahmanian and Srijan Kumar. 2017. Predicting human behavior: The next frontiers. *Science* 355, 6324 (2017), 489–489.

[45] Junting Ye and Leman Akoglu. 2015. Discovering opinion spammer groups by network footprints. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 267–282.

[46] Reza Zafarani and Huan Liu. 2015. 10 bits of surprise: Detecting malicious users with minimum information. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 423–431.