

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

Tutorial: Data-Driven Approaches towards Malicious Behavior Modeling



Meng Jiang
University of Notre Dame



Srijan Kumar Stanford University



Christos Faloutsos
Carnegie Mellon
University



V.S. Subrahmanian University of Maryland, College Park

http://www.meng-jiang.com/tutorial-kdd17.html

Acknowledgement













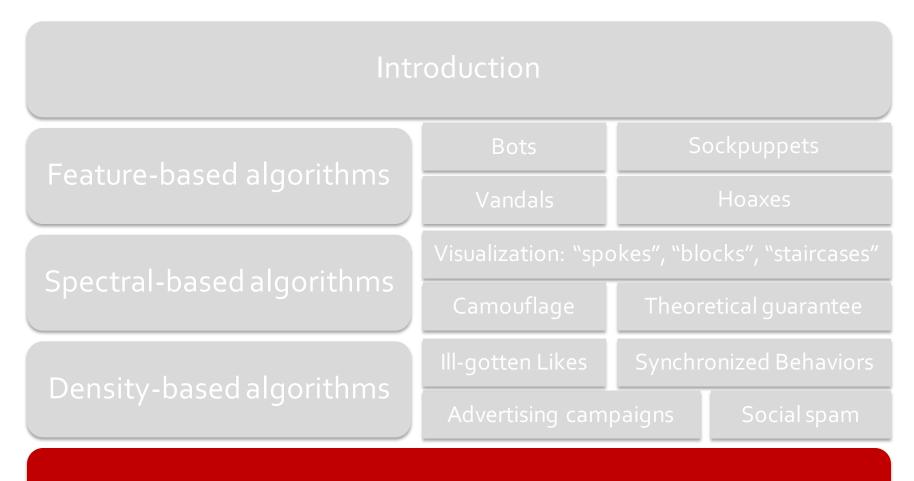








Outline



Conclusions and future directions

Feature based signs of malicious behavior to look out for

- Activity: malicious behavior is often done with "throwaway" and recent accounts
- Temporal: malicious users are often faster
- Linguistic: malicious users are often abusive and more opinionated
- Network: malicious users often collude and are densely connected to each other
- Community feedback: malicious users are harshly treated by other users, but regular negative feedback can be harmful

Summary of each section

ADD

P1. Anonymity

What is the role of anonymity and the lack of single verified identify in antisocial behavior on the internet?

P2. Early detection

How can antisocial behavior and disinformation be detected as early as possible?

What features can we use?
Can we skip semantic analysis and fact checking?

P3. Adversarial setting

Bad users can actively change behavior in presence of new detection measures to avoid detection.

How do we deal with this?

P4. Organized adversaries

How do we detect coordinated attacks on social media?

P5. Multi-platform malicious behavior

How do antisocial entities behave across several platforms?

Datasets

- Wikipedia hoax dataset: www.cs.umd.edu/~srijan/hoax
- Wikipedia personal attack dataset: <u>https://figshare.com/projects/Wikipedia_Talk/16731</u>
- Wikipedia vandals: <u>www.cs.umd.edu/~srijan/vews/</u>
- Wikipedia vandalism: <u>http://wikipapers.referata.com/wiki/List_of_vandalism_dat</u> asets
- TAMU Twitter honeypot dataset: <u>http://infolab.tamu.edu/data/</u>
- Twitter synchronized malicious behavior data: <u>http://www.meng-jiang.com/pubs/catchsync-kdd14/catchsync-kdd14-code-and-data.gz</u>
- Amazon, Yelp, TripAdvisor review datasets:
- http://shebuti.com/collective-opinion-spam-detection/
- http://cs.unm.edu/~aminnich/trueview/
- https://www.cs.uic.edu/~liub/FBS/fake-reviews.html
- http://snap.stanford.edu/data/#reviews



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

Tutorial: Data-Driven Approaches towards Malicious Behavior Modeling



Meng Jiang University of Notre Dame



Srijan Kumar Stanford University



Christos Faloutsos
Carnegie Mellon
University



V.S. Subrahmanian University of Maryland, College Park

http://www.meng-jiang.com/tutorial-kdd17.html