An Introduction to Graph Based Spam Review Detection

Big Data and Social Computing Lab
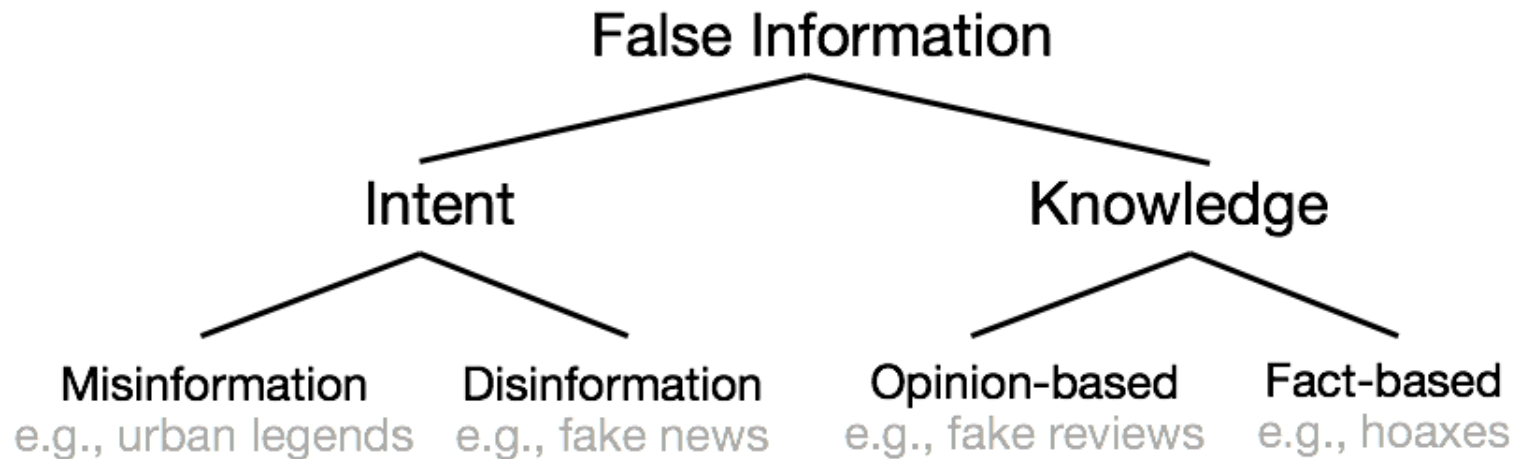**University of Illinois at Chicago**

Yingtong Dou

08/09/2018

Outline

1. Introduction
2. General Methods
3. SpEagle
4. Fraudar
5. FBox
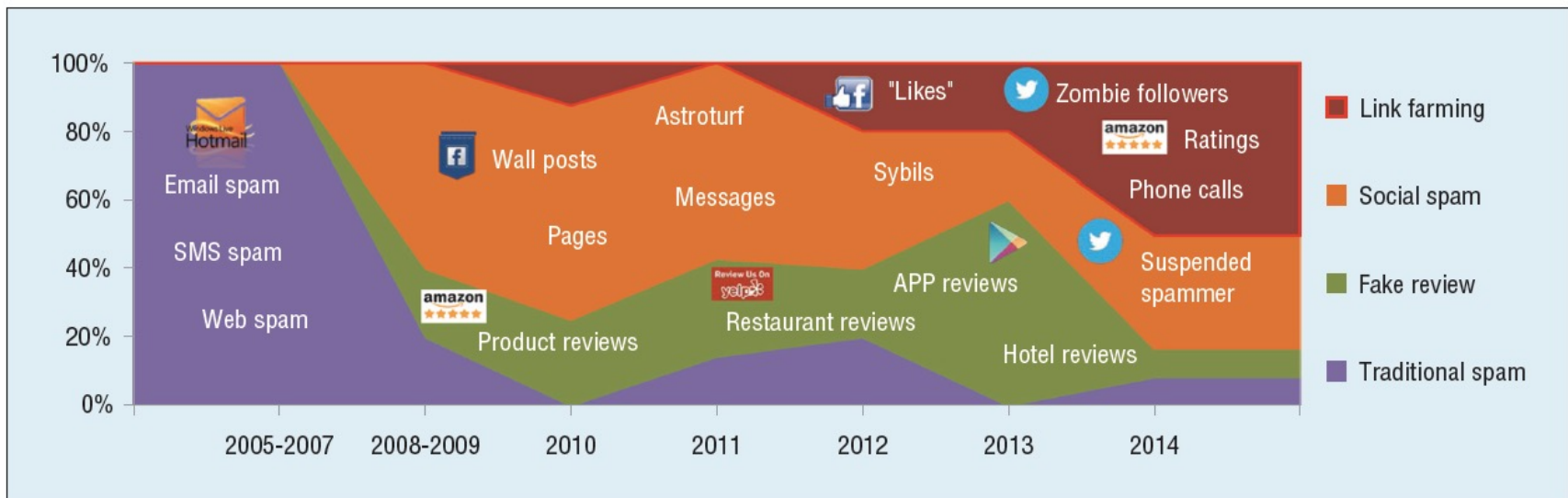6. FairJudge

1. Introduction

Types of false information

Kumar, Srijan, and Neil Shah. "False Information on Web and Social Media: a Survey." *arXiv.org*, April 24, 2018.

1. Introduction

Research Timeline

Jiang, Meng, Peng Cui, and Christos Faloutsos. "Suspicious behavior detection: Current trends and future directions." *IEEE Intelligent Systems* 31.1 (2016): 31-39.

1. Introduction

Examples of fake reviews

Akoglu, Leman, Rishi Chandy, and Christos Faloutsos. "Opinion Fraud Detection in Online Reviews by Network Effects.." *Icwsm*, 2013.

2. General Methods

Current methods adopt three approaches below

- Textual features

- Behavior information

- Network structure

Kumar, Srijan, and Neil Shah. "False Information on Web and Social Media: a Survey." *arXiv.org*, April 24, 2018.

2. General Methods

Textual features

Avg. content similarity

Avg. review length in number of words

Percentage of ALL-capitals words

Percentage of capital letters

Review length in words

Ratio of 1st person pronouns ('I', 'my', etc.)

Ratio of exclamation sentences containing '!'

Ratio of subjective words (by sentiWordNet)

Ratio of objective words (by sentiWordNet)

Description length (information-theoretic)

Rayana, Shebuti, and Leman Akoglu. *Collective Opinion Spam Detection: Bridging Review Networks and Metadata*. *The 21th ACM SIGKDD International Conference*, New York, New York, USA: ACM, 2015.

2. General Methods

Behavior information

Max. number of reviews written in a day
Ratio of positive reviews (4-5 star)
Ratio of negative reviews (1-2 star)
Avg. rating deviation
Weighted rating deviation
Burstiness—spammers are often generate spams in short time.
Absolute rating deviation from product's average rating
Extremity of rating

Rayana, Shebuti, and Leman Akoglu. *Collective Opinion Spam Detection: Bridging Review Networks and Metadata*. *The 21th ACM SIGKDD International Conference*, New York, New York, USA: ACM, 2015.

2. General Methods

Network structure

- SpEagle
- Fraudar
- fbox
- FairJudge

3. SpEagle

Intuition

Akoglu, Leman, Rishi Chandy, and Christos Faloutsos. "Opinion Fraud Detection in Online Reviews by Network Effects.." *Icwsm*, 2013.

3. SpEagle

Belief Propagation at Probabilistic Graph Model

$$m_{X \to Y}(Y) = \sum_{x} \phi_{XY}(x,Y) \prod_{W \in neighbor(X)/Y} m_{W \to X}(x)$$

3. SpEagle

Model Illustration

Rayana, Shebuti, and Leman Akoglu. *Collective Opinion Spam Detection: Bridging Review Networks and Metadata*. *The 21th ACM SIGKDD International Conference*, New York, New York, USA: ACM, 2015.
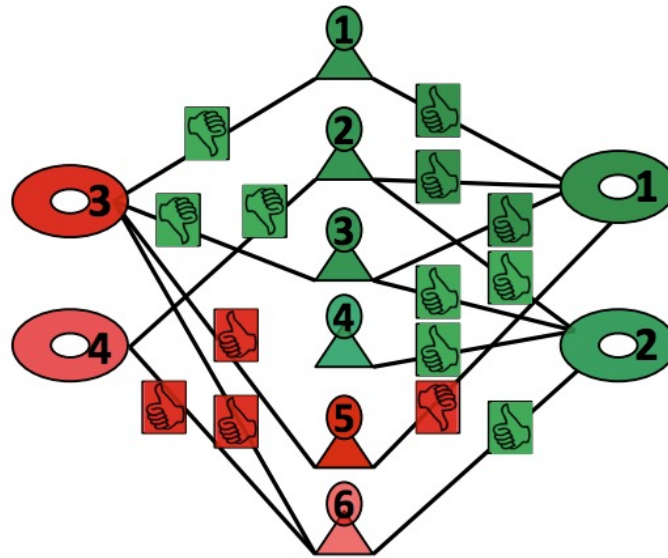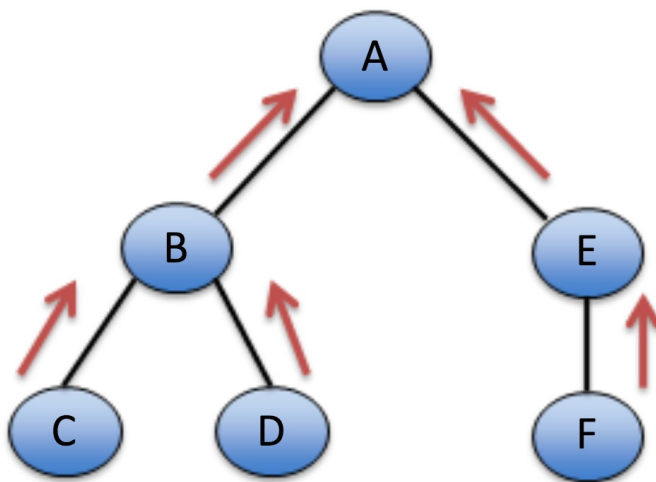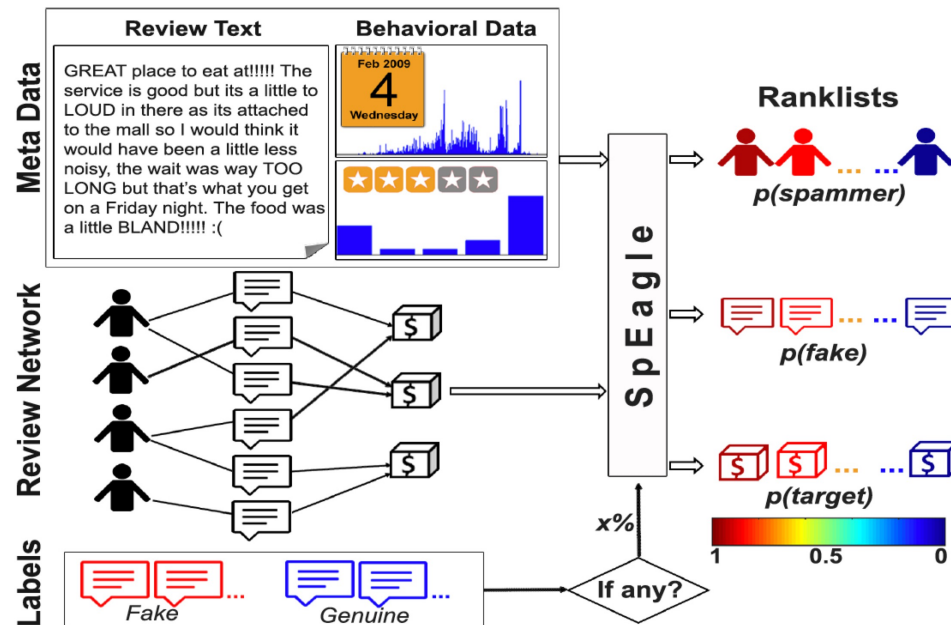
4. Fraudar

- KDD 2016 Best Paper
- Detect dense blocks in the graph
- Camouflage resistant
- Scalability
- Efficient under real world setting
- Theory guarantee

Hooi, Bryan, Hyun Ah Song, Alex Beutel, Neil Shah, Kijung Shin, and Christos Faloutsos. "FRAUDAR - Bounding Graph Fraud in the Face of Camouflage.." *KDD*, 2016, 895–904

4. Fraudar

Algorithm

density = \sum(suspicious score)/number of nodes

Construct priority tree

For node i in graph:

      Calculate density of each graph excluding i

      Find the largest density and store the graph after removing i

      Update priority tree

Find the largest density graph among all stored subgraph and mark them as spammer

Hooi, Bryan, Hyun Ah Song, Alex Beutel, Neil Shah, Kijung Shin, and Christos Faloutsos. "FRAUDAR - Bounding Graph Fraud in the Face of Camouflage.." *KDD*, 2016, 895–904
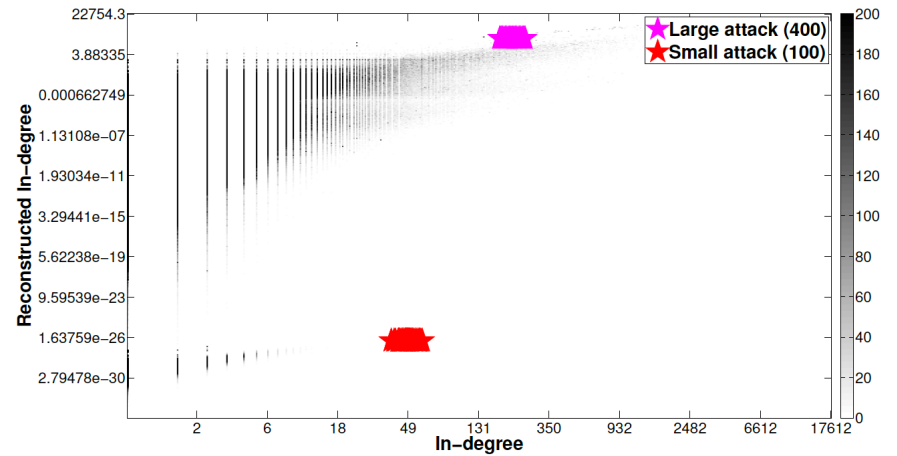
5. fBox

- Attacks under specific threshold will be missed for baseline methods

- Improve Singular Value Decomposition via reconstructing in/out degrees

Shah, Neil, Alex Beutel, Brian Gallagher, and Christos Faloutsos. "Spotting Suspicious Link Behavior with fBox - an Adversarial Perspective.." *Icdm*, 2014, 959–64.

5. fBox
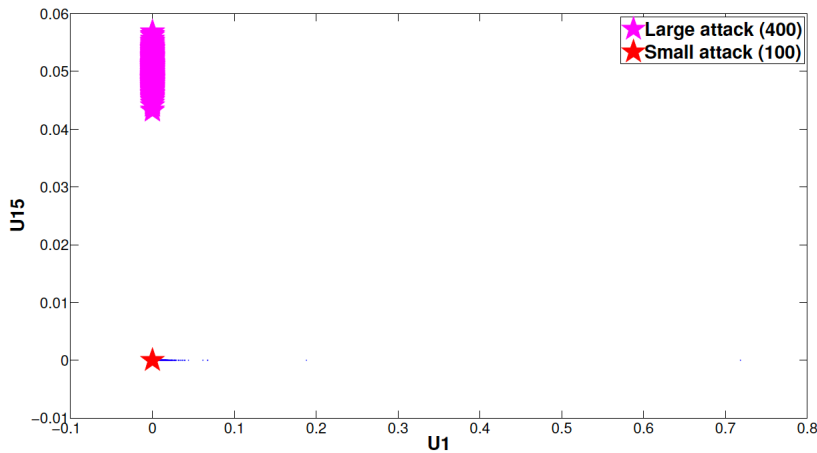
(a) Spectral subspace plot

(b) ISRM

6. FairJudge

- Define three metrics:
  - Fairness of user
  - Reliability of rating
  - Goodness of product
- Iteratively update three scores above
- Provide theoretical guarantee
- Linear complexity

Kumar, Srijan, Bryan Hooi, Disha Makhija, Mohit Kumar, Christos Faloutsos, and V S Subrahmanian. "FairJudge - Trustworthy User Prediction in Rating Platforms.." *CoRR* cs.SI (2017).

Thanks!