

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

Acknowledgement



腾讯
Tencent

Outline

Introduction

Feature-based algorithms

Bots

Sockpuppets

Vandals

Hoaxes

Spectral-based algorithms

Visualization: “spokes”, “blocks”, “staircases”

Camouflage

Theoretical guarantee

Density-based algorithms

Ill-gotten Likes

Synchronized Behaviors

Advertising campaigns

Social spam

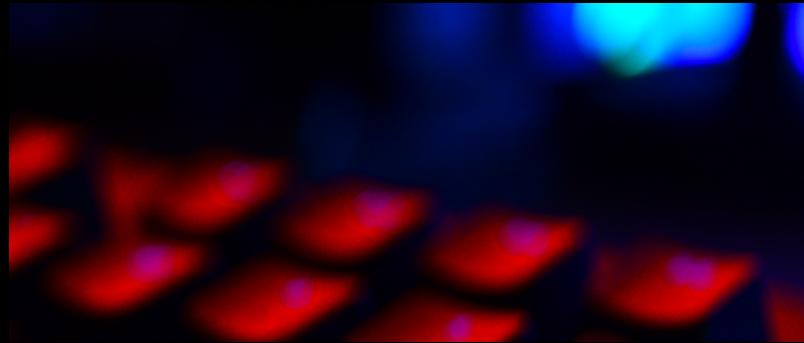
Conclusions and future directions

Web: A platform for everyone



Web allows us...

- Web enables social interaction
- Web is no longer a static library that people passively browse
- Web is a place where people:
 - Act as prosumers, i.e., content producers and content consumers
 - Interact with other people:
 - Internet forums, Blogs, Social networks, Twitter, Wikis, Podcasts, Slide sharing, Bookmark sharing, Product reviews, Comments, ...



...but there's also a dark side
to the web



Not everyone has good intentions...

How Trolls Are Ruining the Internet

Wikipedia blocks hundreds of 'scam' sock puppet accounts

When Will the Internet Be Safe for Women?

FAKE NEWS IS ABOUT TO GET EVEN SCARIER THAN YOU EVER DREAMED

Fake reviews on the Play Store reportedly growing and getting smarter

On the web...

Anyone can use multiple identities
to put something on the internet.

From anywhere in the world.

They can say anything they like.

Leave it there as long as they like.

Change it whenever they feel like it.

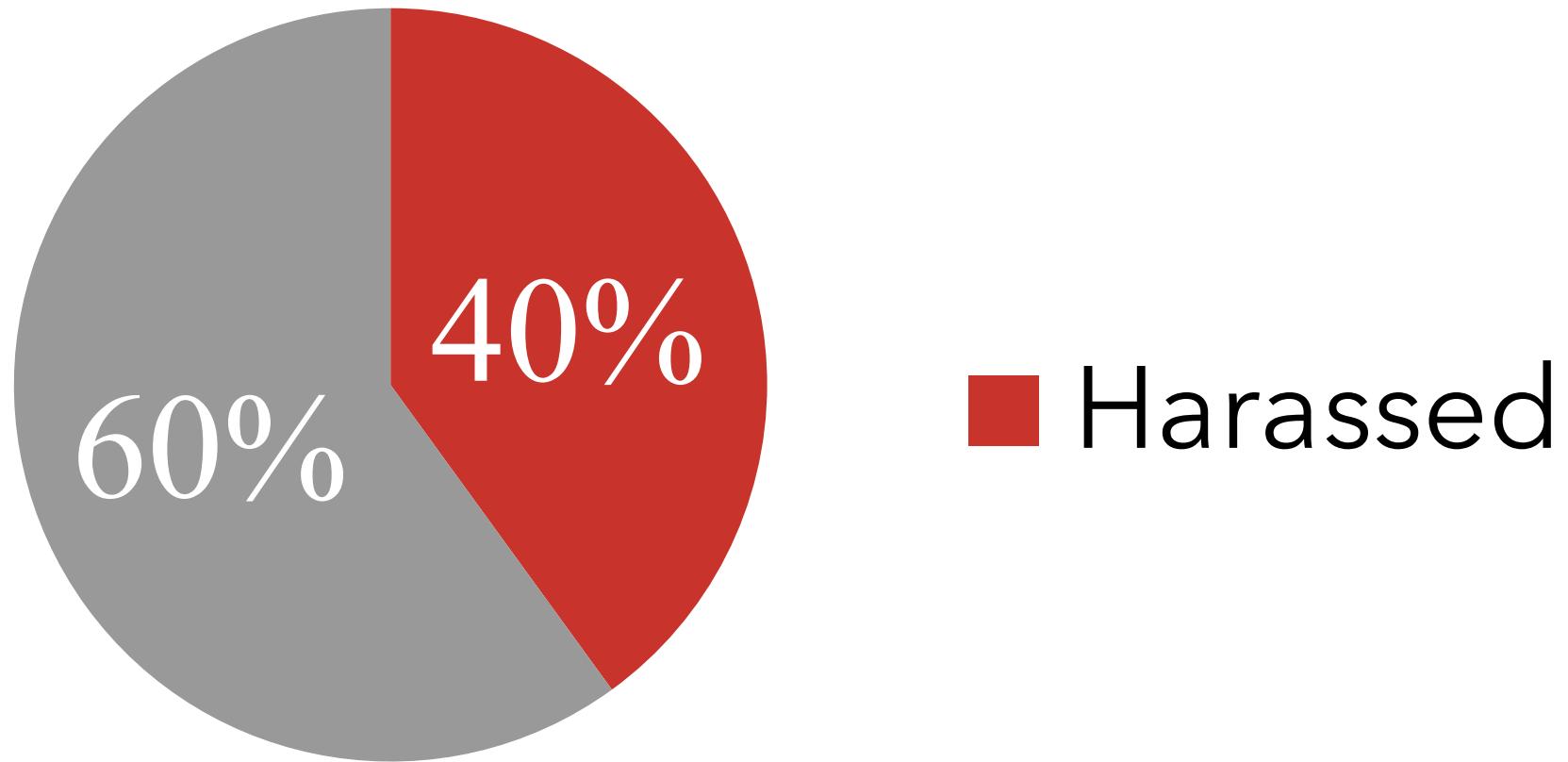
Anonymity on the web exacerbates the problem



"On the Internet, nobody knows you're a dog."

©The New Yorker Collection 1993 Peter Steiner
From cartoonbank.com. All rights reserved.

Malicious behavior is prevalent online



Pew Research (2014)

Malicious behavior has harmful effects



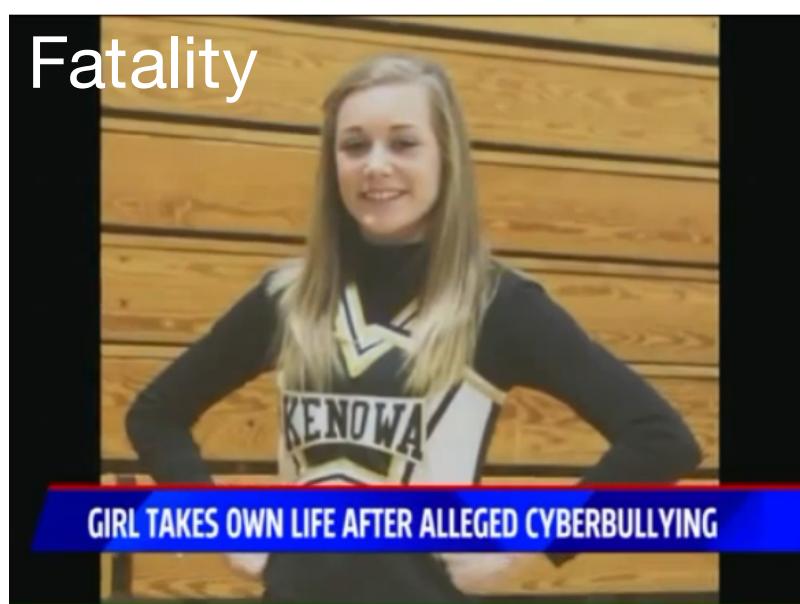
Distress



Harassment



Delinquent behavior



GIRL TAKES OWN LIFE AFTER ALLEGED CYBERBULLYING

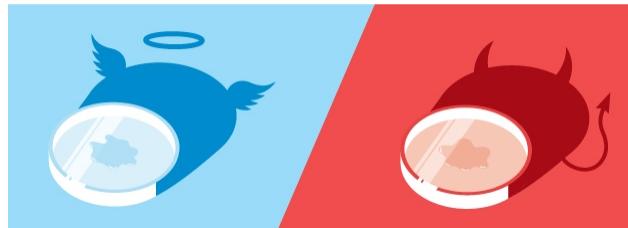
This thesis:
Characterization and Detection of
Malicious Behavior on the Web

Challenges in analyzing malicious behavior



Data imbalance

Smaller proportion of behavior (< 10%) is malicious



Limited labels

Little known information about malicious behavior



Deceptive behavior

Malicious behavior tends to masquerade as benign

Goals of this tutorial

- Discuss a set of methods for analyzing malicious behavior online
- Summarize insights obtained by various studies into different aspects of malicious behavior online:
 - Malicious users
 - Malicious content

Overarching Takeaway

Malicious behavior differs from benign behavior both with respect to their attributes and connectivity.

These differences can be used to create efficient computational methods to detect them.



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