Tracking, Revealing and Detecting Crowdsourced Manipulation

Kyumin Lee, Utah State University
http://digital.cs.usu.edu/~kyumin/cm

The objective of this project is to

- Detecting Malicious Tasks in Crowdsourcing Platforms and Building a Task Blacklist
- Revealing the Ecosystem of Crowdturfers and Detecting Crowdturfers in Target Sites
- Comparing and Detecting Crowdturfers and Other Malicious Participants

Examples of Malicious Tasks

Facebook Like: Page

Work done: **19/190**
You will earn **$0.45**
Task takes less than **2** min to finish
Job ID: d60eb6a5f88d

Employer: Member_844954
add to Exclude List
add to Include List
Tasks will be rated within **7** days

Facebook → Facebook Like (50+ Friends)

What is expected from Workers?

You must have 50+ Facebook friends to do this task.

1. Go to https://www.facebook.com/
2. Search for "Manhattan-Cardiovascular-Associates"
2. "Like" the page

Required proof that task was finished?

1. Your Facebook display name
2. URL to your Facebook profile
Make sure you have set your profile to Public View in order for your task to be verified.

3. Number of Facebook friends on your account

Post a review >>> Earn 10 Cents >>> Fast Acceptance

Work done: 8/20
You will earn: $0.10
This task takes less than 3 minutes to finish
Campaign ID : 57aabd71-5e3c-48c9-acdc-e4e93257911a
Campaign Name : Post a review >>> Earn 10 Cents >>> Fast Acceptance

You can accept this job if you are from THESE COUNTRIES ONLY:

International

What is expected from workers?

1. Signin to your Amazon Account. 2. Go to goo.gl/Eirdfu 3. Download the book and leave a review. 4. Review should be in correct Grammar and complete sentences. 5. Review must be 40 words long.

Required proof that task was finished?

1. Link to your review
2. Your Reviewer id
3. Reviews posted by Anonymous id will not be accepted

Project Results So far…

Detecting Malicious Campaigns in Crowdsourcing Platforms

Dataset

- Definition of malicious campaigns: require workers to manipulate information in targeted sites
- Collected 23,220 campaign descriptions from four crowdsourcing platforms between November 2014 to January 2015
  - Mechanical Turk, Microworkers, Rapidworkers, and Shorttask

Market size and Hourly wages

	Malicious	Legitimate
# of Tasks	798,796 (23.8%)	2,557,357 (76.2%)
Market Size	$148,911	$179,696
Hourly Wage	$2.48	$1.88

- The malicious tasks (24%) occupied 45% of the entire market
- Interestingly, hourly wage for malicious tasks was much higher than one for legitimate tasks
- Malicious campaigns mostly targeted social networking sites (60%) and search engines (30%)

Features

- Numerical features
  - Reward, # of tasks, estimated time to complete, hourly wage, # of URLs in task instruction, # of words and # of words in a task title, ratio of # of URLs to # of words
- Text features
  - Extracted unigram, bigram and trigram features from task title and instruction

Detecting Malicious Campaigns

- 3 baseline methods
  - Majority selection (i.e., legitimate campaign)
  - URL-based filtering: malicious if it contains top K sites' URLs
  - Principal Component Analysis (PCA) [Viswanath et al. USENIX Security 2014]

Approach	Accuracy	FPR	FNR
Majority Selection	78.4%	1	0
URL-based filtering@100	72.4%	0.708	0.157
PCA – 12% threshold	85.2%	0.999	0.031
our Naïve Bayes	89.0%	0.044	0.147
our J48	99.1%	0.023	0.058
our SVM	99.2%	0.019	0.055

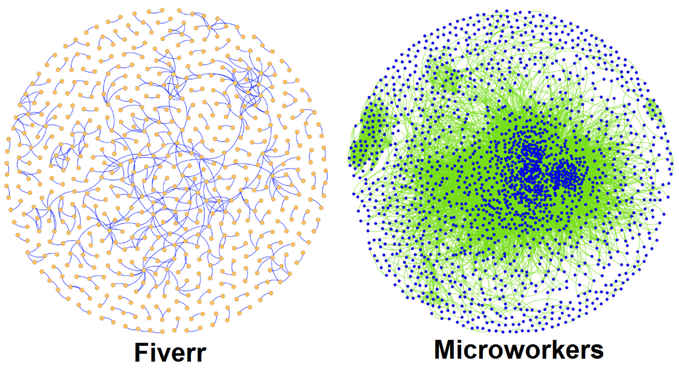
Uncovering Fake Likers in Online Social Networks

Data Collection

- Deployed honeypots to Fiverr and collected 3,207 fake Likers' profiles
- Collected 3,688 fake Likers' profiles by linking *Like* tasks in Microworkers
- Collected 6,252 legitimate Likers' profiles from conference groups and random pool with manual verification

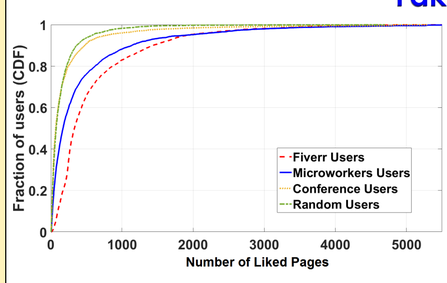
Understanding Fake Likers

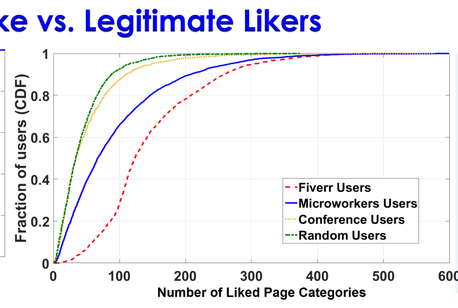
- Delivered fake likes in a bursty fashion
- More males participated in the fake liking activity.
- 73% Likers were from top 10 countries – mostly developing countries.
- Most fake likers were in a range of 18-34 years old



Mutual friendship relations in Fiverr and Microworkers communities

Fake vs. Legitimate Likers





Fake Likers…
- performed more page liking activities
- liked more diverse categories
- had shorter longevity

Detecting Fake Likers

Top 5 Features	Fake Likers (Avg. Value)	Legitimate Likers
Category Entropy	6.35	4.47
Longevity	3.62	5.53
Average # of posts per day	0.31	0.12
# of lines in *About* section	4.02	4.08
Proportion of verified pages	0.23	0.26

Approach	Accuracy	FPR	FNR
PCA	0.690	0.30	0.32
SynchroTrap	0.505	0	0.99
CopyCatch	0.565	0.85	0.02
PCA - Test	0.690	0.30	0.32
SynchroTrap - Test	0.635	0	0.73
CopyCatch - Test	0.655	0.46	0.23
our LogitBoost	0.875	0.11	0.13
our Random Forest	0.885	0.09	0.13
our XGBoost	**0.897**	0.08	0.11

- Our XGBoost model outperformed PCA, SynchroTrap and CopyCatch, with 0.897 accuracy in the dataset

Interested in meeting the PIs? Attach post-it note below!

National Science Foundation
WHERE DISCOVERIES BEGIN

The 3rd NSF Secure and Trustworthy Cyberspace Principal Investigator Meeting
January 9-11, 2017
Arlington, Virginia

UtahState University