Identifying Video Spammers in Online Social Networks

Fabrício Benevenuto[1], Tiago Rodrigues[1], Virgílio Almeida[1], Jussara Almeida[1], Chao Zhang[2], Keith Ross[2]

[1] **Federal University of Minas Gerais – Brazil**

[2] **Polytechnic University – New York, USA**

International Workshop on Adversarial Information Retrieval on the Web (AirWeb'08)
Beijim, China April 22, 2008

Motivation

- Video as new trend
  - including political debates, video chats, video mail, and video blogs
- Web services offers video-based features as alternative to text-based
  - video reviews for products, video ads, video responses
  - Susceptible to different types of malicious and opportunistic user actions
- Video response feature: video sequence that begins with an opening video and then followed with video responses
  - Video response spam is a video posted as a response, but whose content is completely unrelated to the opening video.
  - Possible reasons for video response spam:
    - increase the popularity of a video, marketing advertisements, distribute pornography, or simply pollute the system

Example of video response spam

Video

Flintstones - Happy Anniversary

Video Response Spam

Sexy Teen Dance

- Video pornography posted as video response to a cartoon

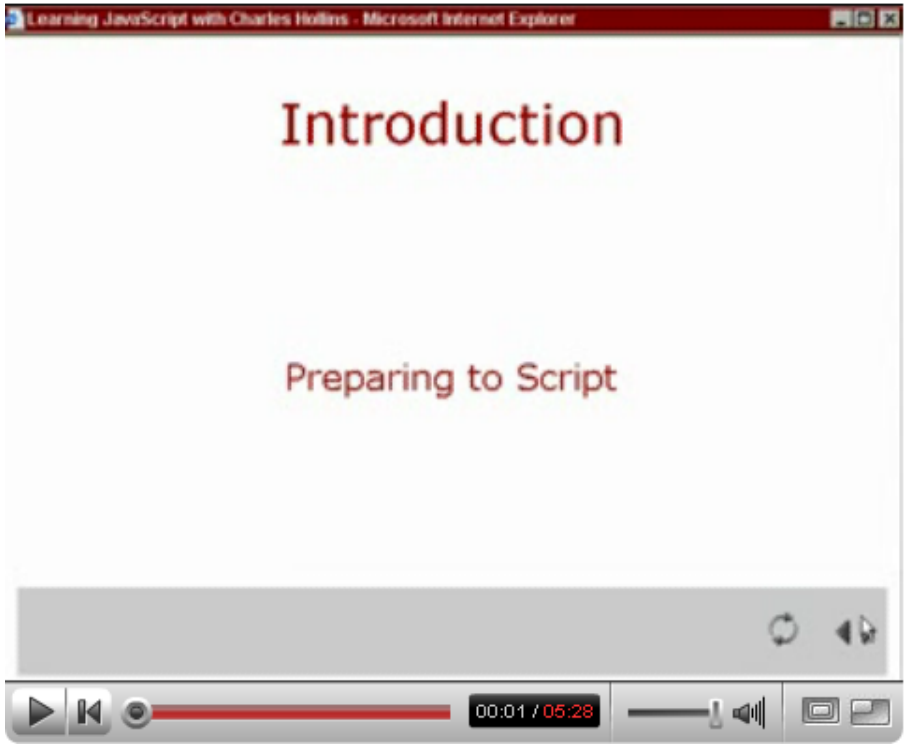
Example of video response spam

Video

Miss Teen USA 2007 - South Carolina answers a question

Video Response Spam

Learn Javascript (Lynda.com) chapter1 -partsix (1/2)



- Advertising of Lynda.com, teaching to program on Javascript as a video response to a very popular video of Miss in troubles to answer a question

Example of video response spam

Video

Liverpool 4 - 2 Arsenal Uefa Champions League

Video Response Spam

Free Web Proxy - Air-Proxy.com

- Advertisement of a proxy service as video response to a soccer game video: Liverpool x Arsenal

Goals

- Quantify the evidence of video spamming activity
  - Approach: identify spammers instead of video spam
- Identify attributes able to distinguish spammers from legitimate users
- "Manually" create a test collection of spammers and legitimate users on YouTube
  - Challenge: the definition of video spam is subjective
- Propose a mechanism to detect video spammers based on the attributes identified

Sampling video responses

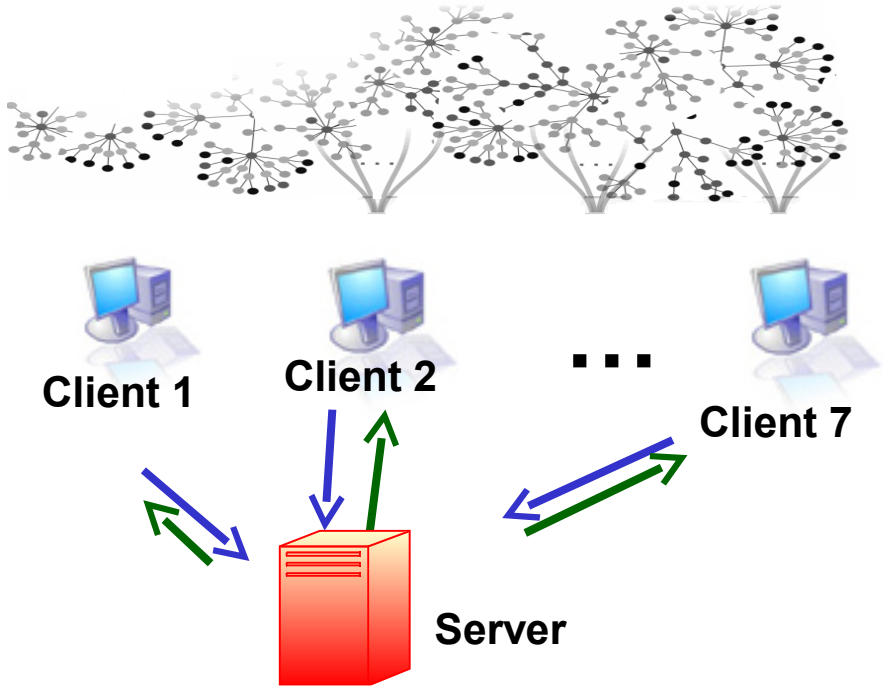
- Vide Response user graph


```
input : A list of users (seeds)
1.1  foreach User U in the crawler list do
1.2      Collect U's info using the YouTube API;
1.3      Collect U's video list using the API;
1.4      foreach Video V in the video list do
1.5          Copy the HTML of V;
1.6          if V is a responded video then
1.7              Copy the HTML of V's video
                 responses;
1.8              Insert the responsive users in the
                 crawler list;
1.9          end
1.10         if V is a video response then
1.11             Insert the responded user in the
                 crawler list;
1.12         end
1.13     end
1.14 end
```

- Approach: Collect an entire weak connected component
  - Follow both directions: video responses and video responded
  - For each user **U**, collect all his video responses and video responded. The owners of the videos responded by u and the owners of the videos responses posted to **U**'s videos are added to the crawler
  - This approach allow us to use several social network metrics

Crawler Architecture

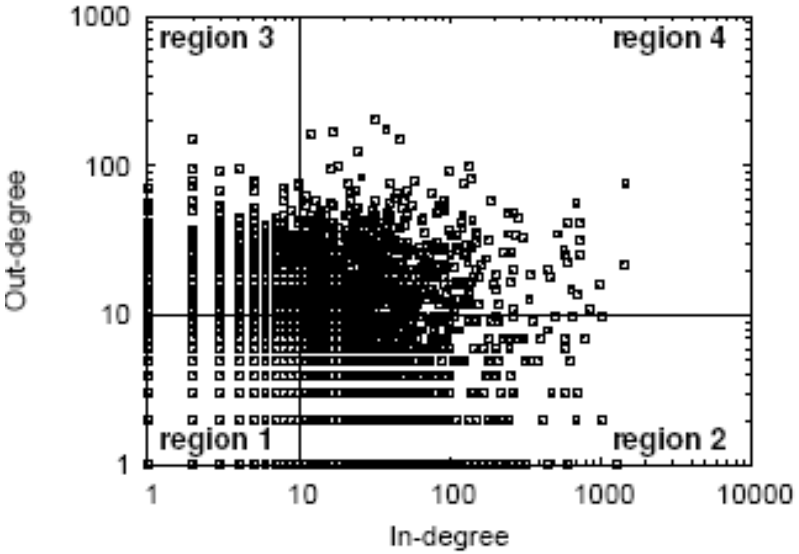
- Clients collects YouTube data
- Server coordinates clients to avoid redundant data collection
- Seeds: users owners of videos of the 100 top responded list



- Collected information of 701,950 video responses and 381,616 responded videos, exhausting an entire component of 264,460 users in 7 days (from Jan 11th to 18th, 2008)

Test Collection

1) Users with different levels of interaction through video responses
   - Select users from 4 different regions of a graph of in-degree x out-degree.
   - Select 100 users from each region. 381 legitimate and 11 spammers (8 with account closed or suspended)

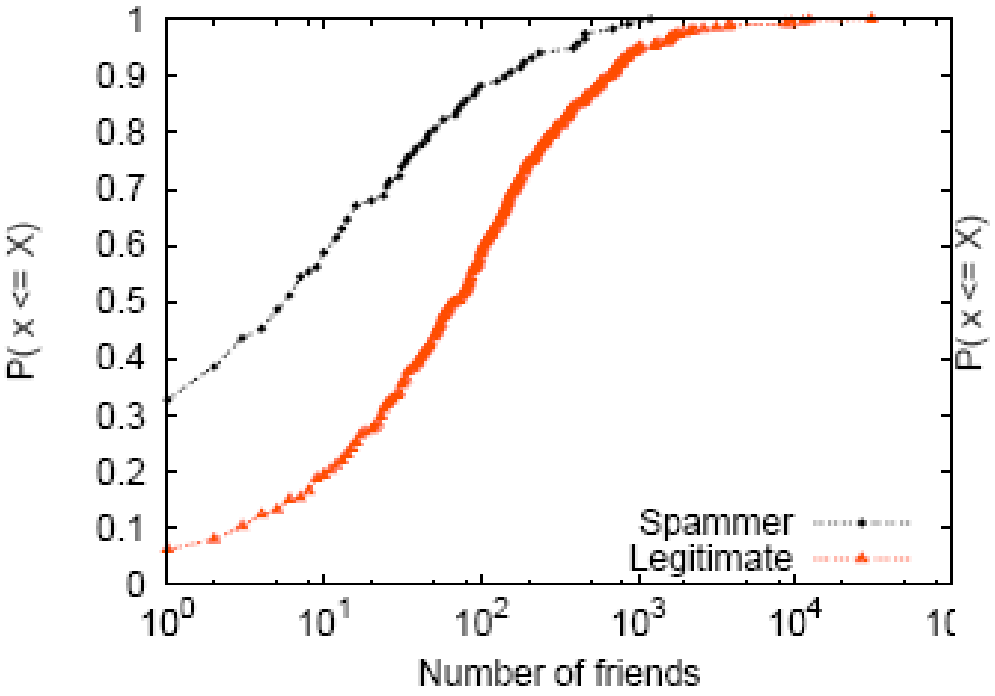


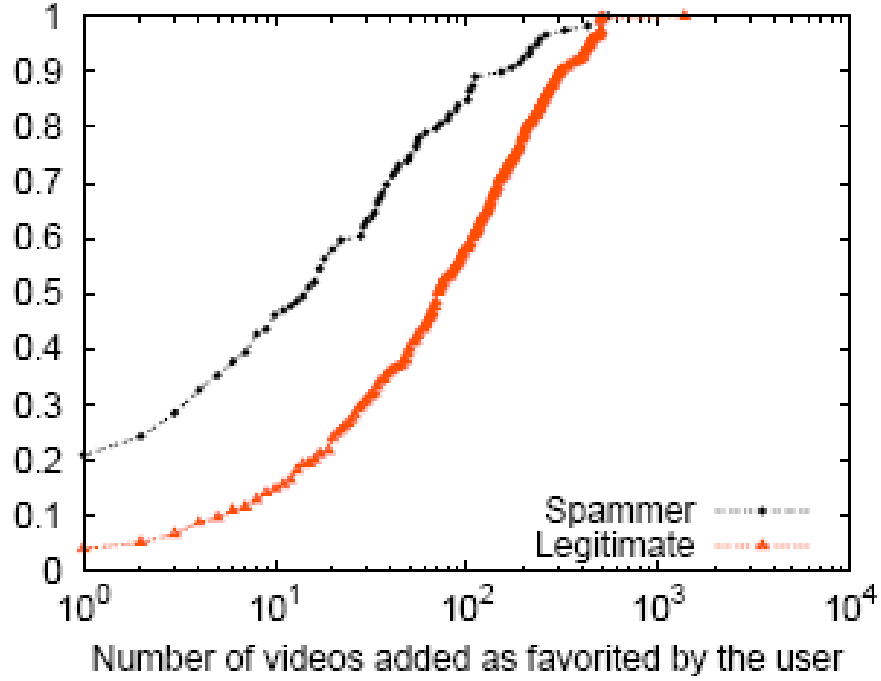
2) Randomly select 100 users from those who posted video responses to videos in the top 100 most
   - 92 legitimate users and 8 spammers

3) Identification of spammers by analyzing the thumbnails of the video responses posted to videos occupying top positions in the top 100 most responded ranking kept by YouTube
   - 100 spammers

- **TOTAL: 592 users, 473 legitimate and 119 spammers**

Characteristics of User profile

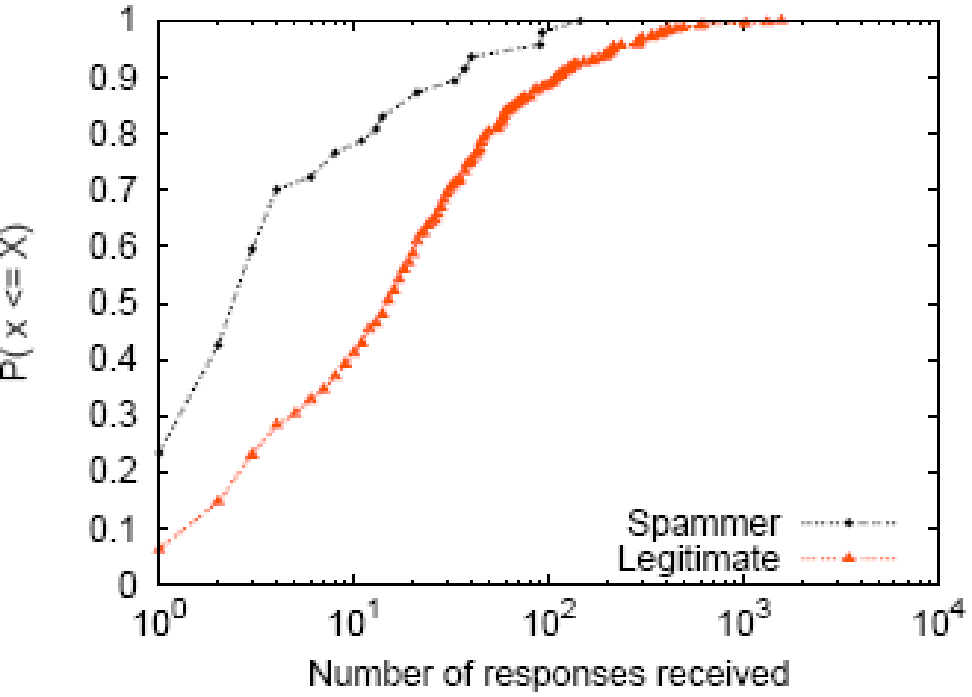


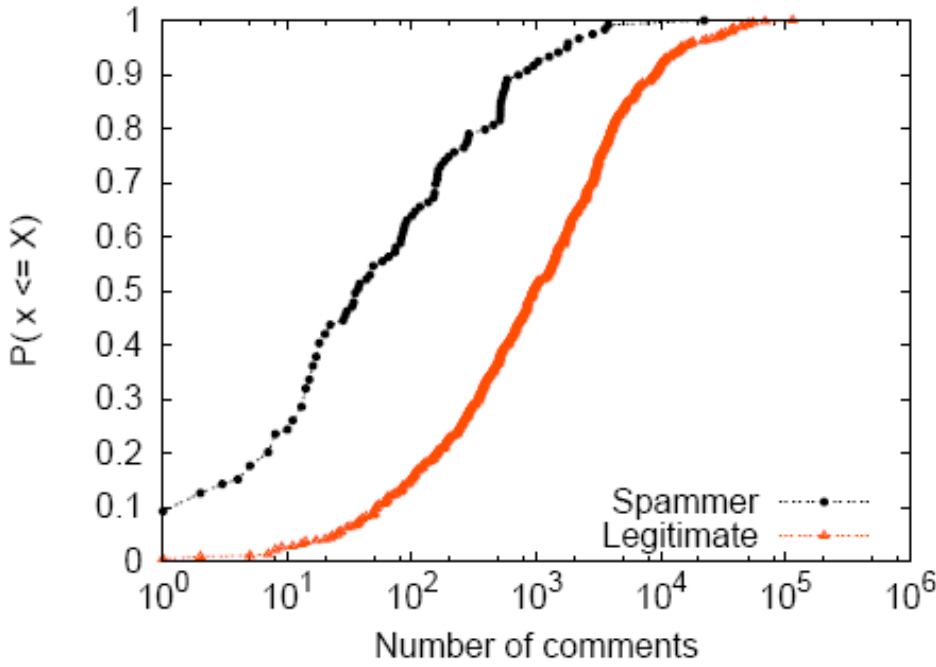


- Legitimate users exhibit a higher level of interaction with the system.
  - Eg. 19% of the legitimate users have less than 10 friends while 56% of the spammers have less than 10 friends.

Characteristics of Videos

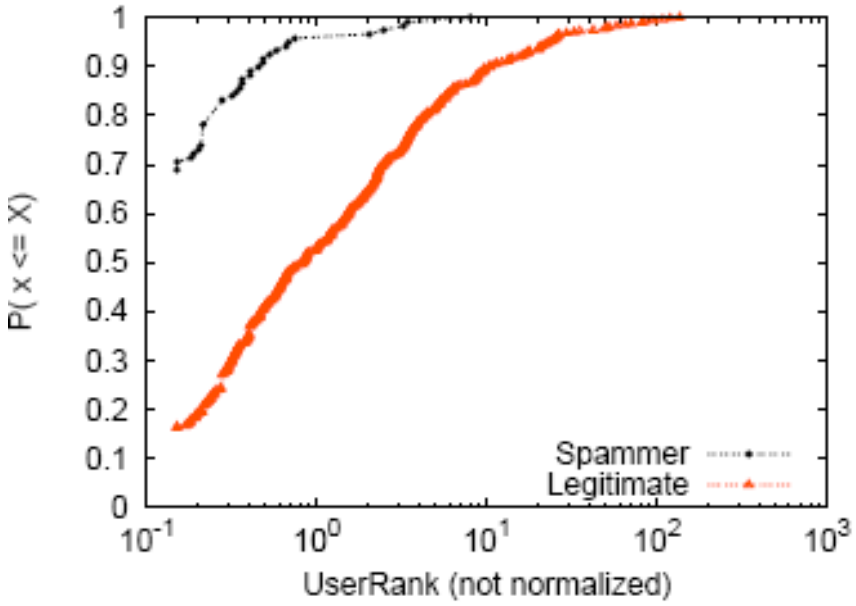
- Quality of the contributions made by users
  - Eg. number of video responses and comments received
  - Characteristics of all videos and only video responses

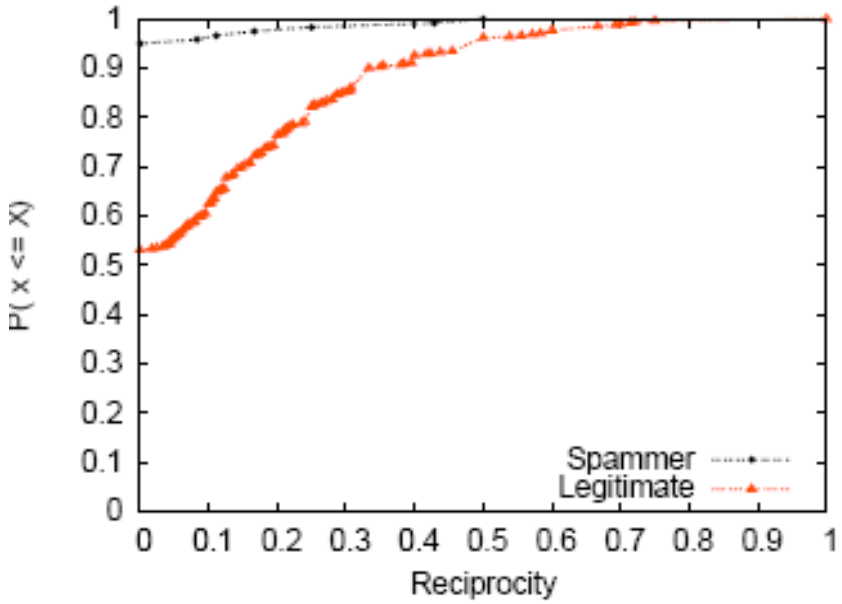




- Plots reflect how other users "view" the quality of the contributions of the two classes of users

Social Network characteristics





- Reciprocity: probability of a user receiving a video response from each user he/she sent a video response.
  - Spammers basically don't have reciprocal links
- UserRank: pagerank algorithm applied on the video response user graph.
  - Importance of the user in terms of his participation on interactions
  - Legitimate users, in general, have a higher UserRank than spammers

Spam detection Mechanism

- Metrics
  - True Positive (TP) , True Negative (TN), False Positive (FP), False Negative (FN), Accuracy, and F-measure
- Features
  - User-Based Features: number of videos uploaded, the number of friends, number of videos watched, number of videos added as favorites, number of video responses posted, number of video responses received, number of subscriptions, number of subscribers
  - Video-Based Features:
    - Average and total for each attribute for 2 groups of videos: all videos of the user and only the video responses.
    - number of views, duration, number of ratings, number of comments, number of favorites, number of honors, number of external links
  - Social Network Features: node in-degree, out-degree, clustering coefficient, UserRank, betweenness, reciprocity, and assortativity

Spam detection Mechanism

- Used SVM (Support vector machine) as classifier
  - 5-fold cross validation
  - libSVM, which allows searching for the best classifier parameters

Metric	User	Video	SN	ALL
TP	0.054	0.426	0.375	0.439
TN	0.998	0.922	1	0.981
FP	0.002	0.078	0	0.019
FN	0.946	0.574	0.625	0.561
Accuracy	0.821	0.821	0.874	0.870
F-measure	0.094	0.484	0.540	0.558

- 44% of the spammers are correctly classified as spammers
- 2% of legitimate users classified as spammers
- Video and social network attributes are the most relevant

Attributes Importance

Position	χ^2	Information Gain	Symetrical Uncert
1	Out-degree	Out-degree	Out-degree
2	# comments total (all videos)	PageRank	# responses created
3	Duration total (all videos)	# comments total (all videos)	Duration mean (all videos)
4	# comments mean (video responses)	In-degree	In-degree
5	PageRank	# comments mean (video responses)	# ratings total (all videos)
6	In-degree	Duration total (all videos)	# comments total (all video)
7	Duration mean (all videos)	# responses received	PageRank
8	# comments mean (all videos)	# comments mean (all videos)	Duration total (all videos)
9	# ratings total (all videos)	# ratings total (all videos)	# Comments mean (video responses)
10	# responses received	duration mean (all videos)	Comments mean (all videos)

- Three feature selection methods
  - Chi Squared, Information Gain, and Symmetrical Uncert.
  - From the 10 most important features we have 9 attributes in common, 6 of video-based attributes and 3 social network attributes

Conclusions and Future Work

- In this work we studied the video spam problem in a popular online social, namely YouTube
- Main Contributions
  - Quantitative evidence of video spamming activity in social online video sharing systems, particularly YouTube.
  - identification and characterization of a set of user and video attributes that can be used to distinguish video spammers from legitimate users
  - A test collection of users from YouTube, classified as spammers or legitimate users.
  - A video spammer detection mechanism based on a classification algorithm, which showed to produce reasonably good results
- Future Work
  - Improve classification
  - Consider multi-class to label users (light spammer, heavy spammer)
  - Extend test collection

Questions?

fabricio@dcc.ufmg.br