# Detecting Spammers and Content Promoters in Online Video Social Networks

Fabrício Benevenuto

(joint work with Tiago Rodrigues, Virgílio Almeida, Jussara Almeida, and Marcos Gonçalves)

Federal University of Minas Gerais - Brazil

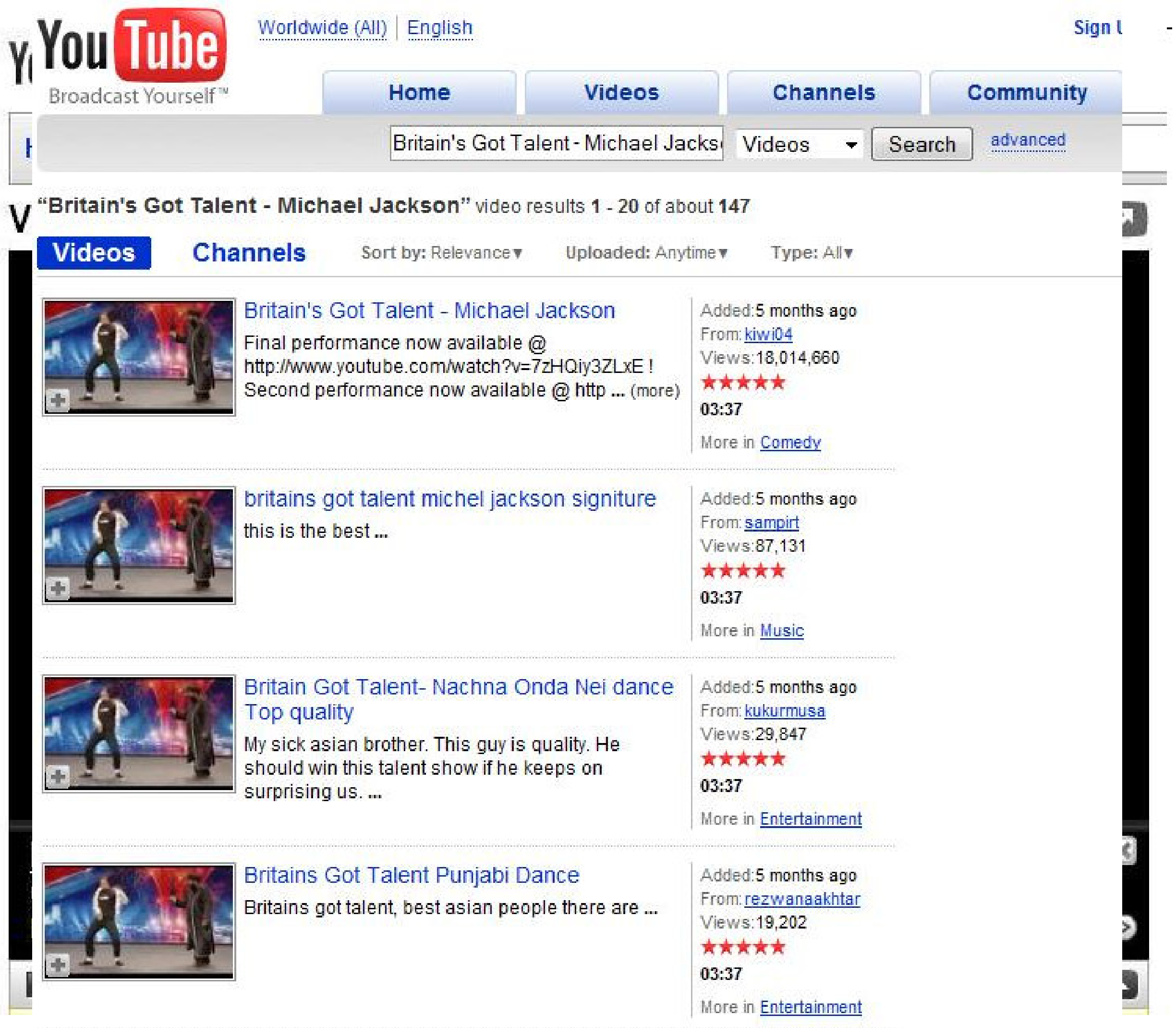
# User Generated Videos

- Video is a trend on the Web
  - YouTube, Yahoo! videos, etc.
  - New features: video review, video blog, video advertises
  - 77% of the U.S. Internet audience viewed online videos
- Explosion of user generated content
  - YouTube has 10 hours of videos uploaded every minute

Users are not only viewing a lot of videos, but they are also creating a lot of videos

# New problems and challenges

- Content retrieval
  - Bad assignment of metadata
  - Duplicates
- System design and infrastructure
- Advertisements
  - The contextual analysis is hard to do
- Opportunistic user actions



# This Talk

Detect opportunistic actions in the YouTube video response feature


Users intentionally post unrelated videos to the video topic

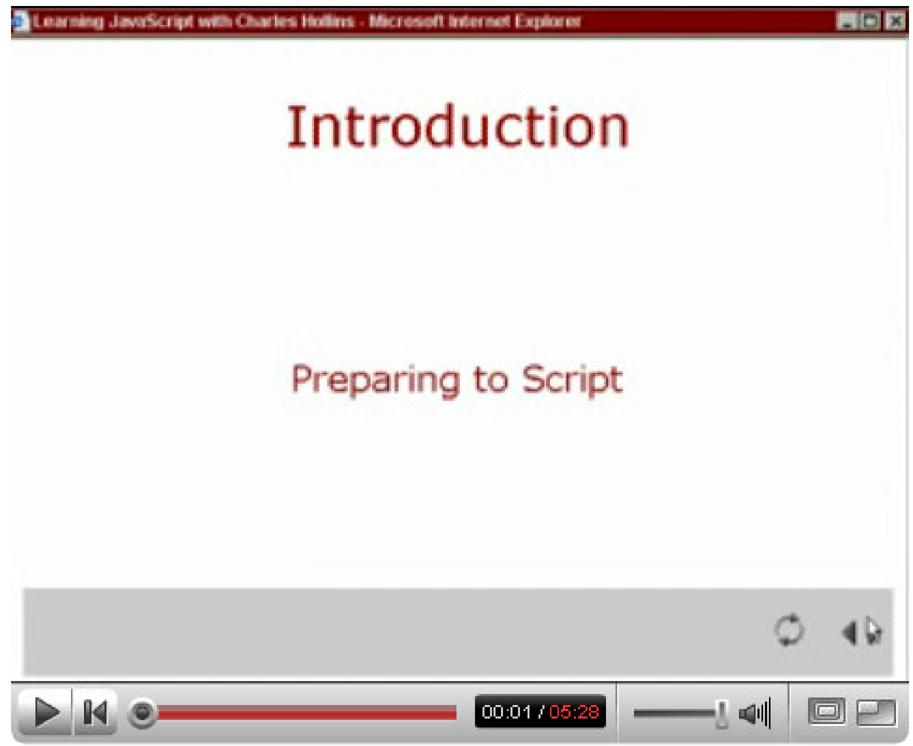
# Example of unrelated videos

## Video

Miss Teen USA 2007 - South Carolina answers a question


## Video response

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 unrelated video

Video

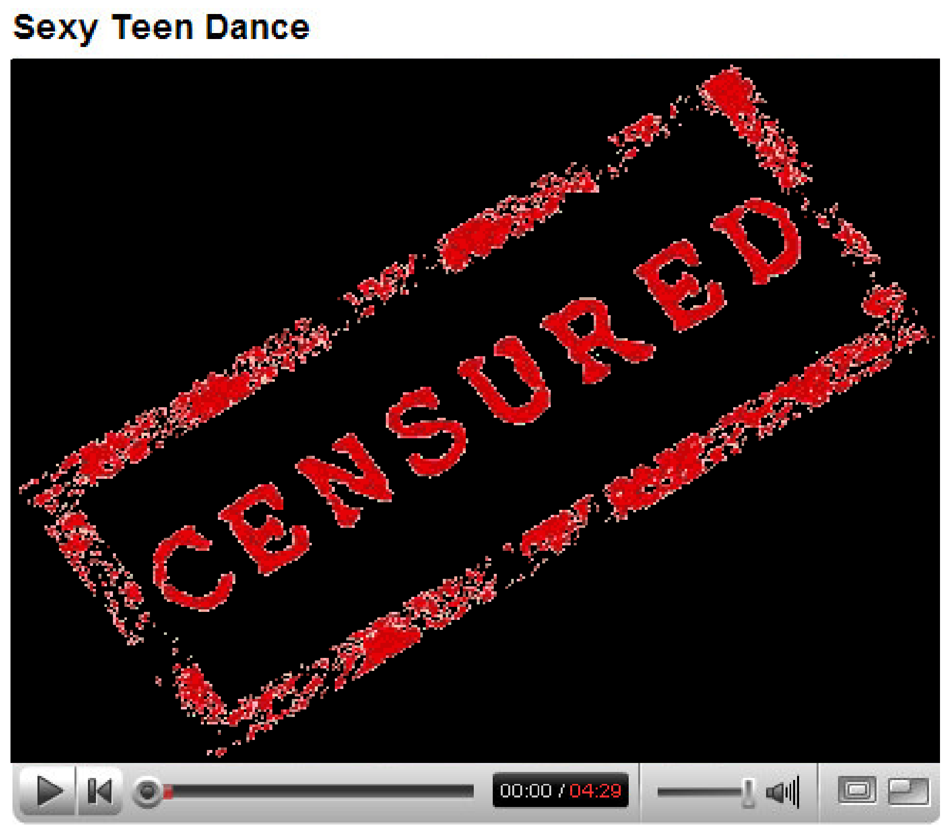

Video Response


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

# Example of unrelated videos

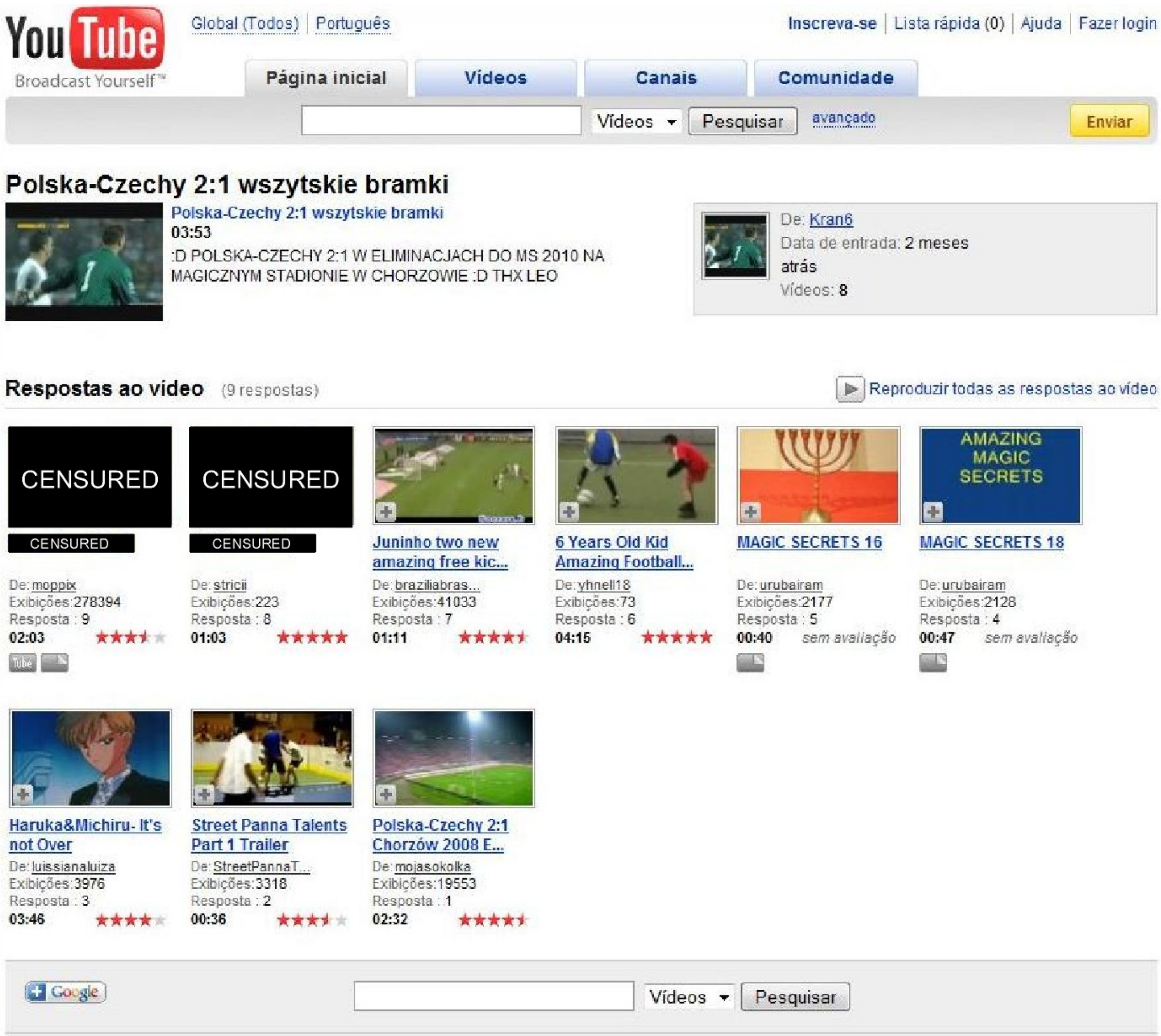
Video

Video response

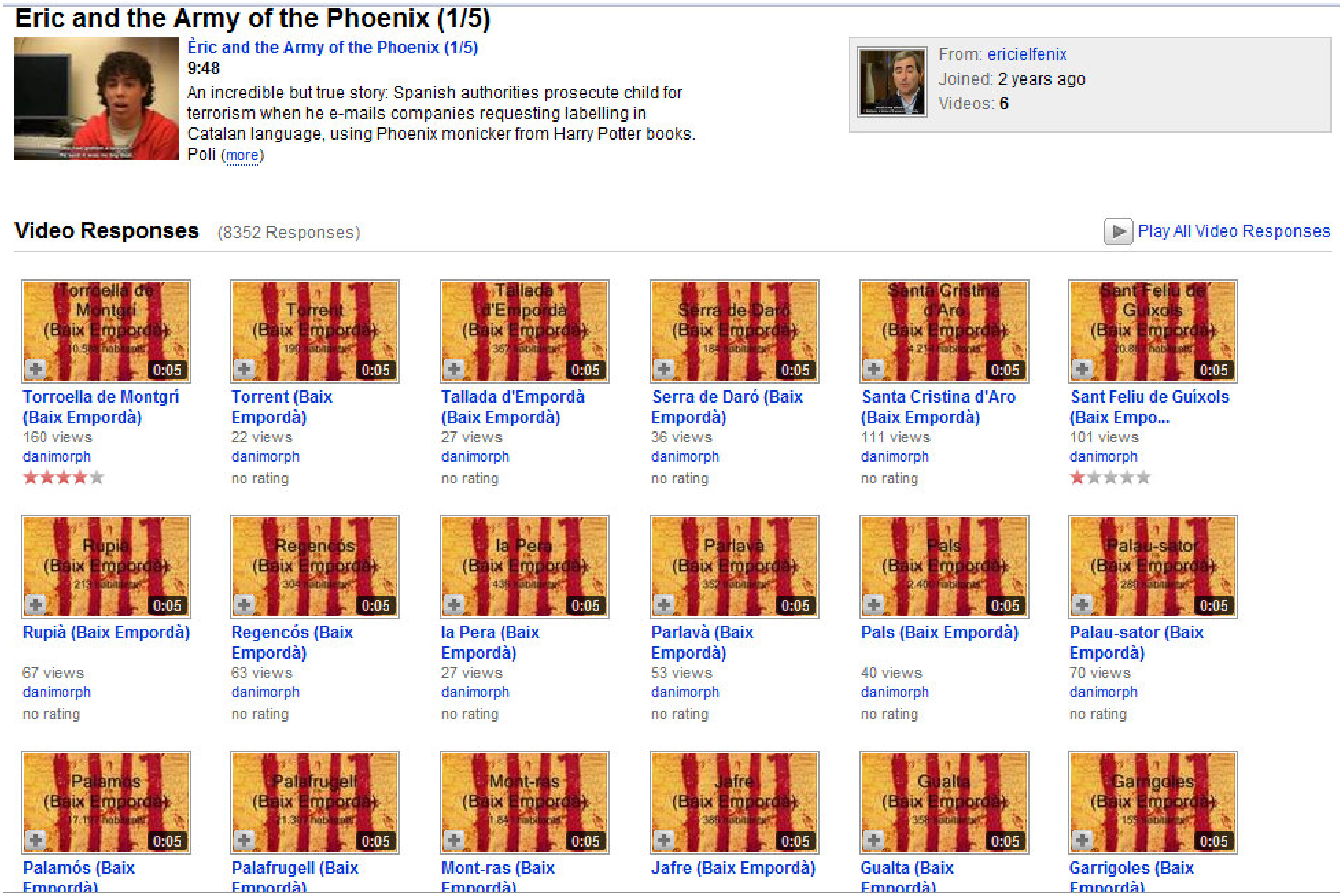




- Video pornography posted as video response to a cartoon

# Video Spam


You Tube
Broadcast Yourself™
Global (Todos) | Português
Inscreva-se | Lista rápida (0) | Ajuda | Fazer login
Página inicial
Vídeos
Canais
Comunidade
Vídeos
Pesquisar
avançado
Enviar
Polska-Czechy 2:1 wszytskie bramki
Polska-Czechy 2:1 wszytskie bramki
03:53
:D POLSKA-CZECHY 2:1 W ELIMINACJACH DO MS 2010 NA MAGICZNYM STADIONIE W CHORZOWIE :D THX LEO
De: Kran6
Data de entrada: 2 meses atrás
Vídeos: 8
Respostas ao vídeo (9 respostas)
Reproduzir todas as respostas ao vídeo
CENSURED
CENSURED
De: moppix
Exibições:278394
Resposta : 9
02:03
CENSURED
CENSURED
De: stricii
Exibições:223
Resposta : 8
01:03
Juninho two new amazing free kic...
De: braziliabras...
Exibições:41033
Resposta : 7
01:11
6 Years Old Kid Amazing Football...
De: yhnell18
Exibições:73
Resposta : 6
04:15
MAGIC SECRETS 16
De: urubairam
Exibições:2177
Resposta : 5
00:40 sem avaliação
AMAZING MAGIC SECRETS
MAGIC SECRETS 18
De: urubairam
Exibições:2128
Resposta : 4
00:47 sem avaliação
Haruka&Michiru- It's not Over
De: luissianaluiza
Exibições:3976
Resposta : 3
03:46
Street Panna Talents Part 1 Trailer
De: StreetPannaT...
Exibições:3318
Resposta : 2
00:36
Polska-Czechy 2:1 Chorzów 2008 E...
De: mojasokolka
Exibições:19553
Resposta : 1
02:32
Google
Vídeos
Pesquisar

# Video Promotion



# Negative Impact of Promotion and Spam

- Challenges for users in identifying video promotion and spam
  - consumes system resources, especially bandwidth
  - compromise user patience and satisfaction with the system
- Pollution in top lists
- Difficulty in ranking and recommendation
  - Promoted or spam videos may be temporarily ranked high or considered related to the video topic

# Goal

- Detect video spammers and promoters

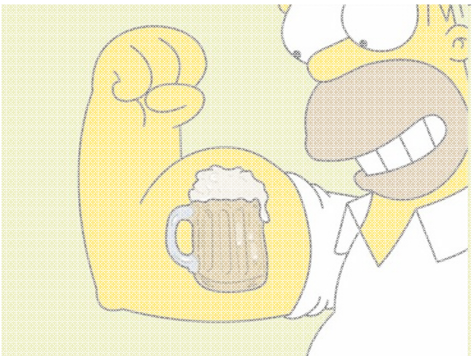
- 4-step approach

1. Sample YouTube video responses and users

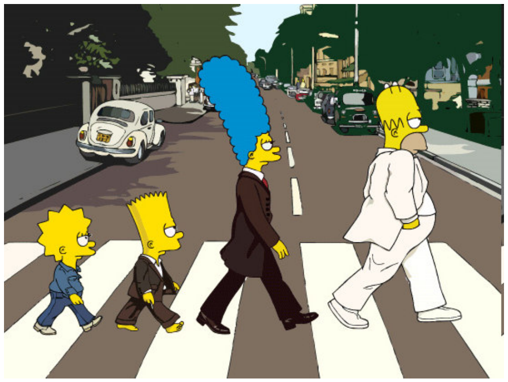
2. Manually create a user test collection
   (promoters, spammers, and legitimate users)

3. Identify attributes that can distinguish spammers and promoters from legitimate users

4. Classification approach to detect spammers and promoters



Part1.
Motivation
& Problem

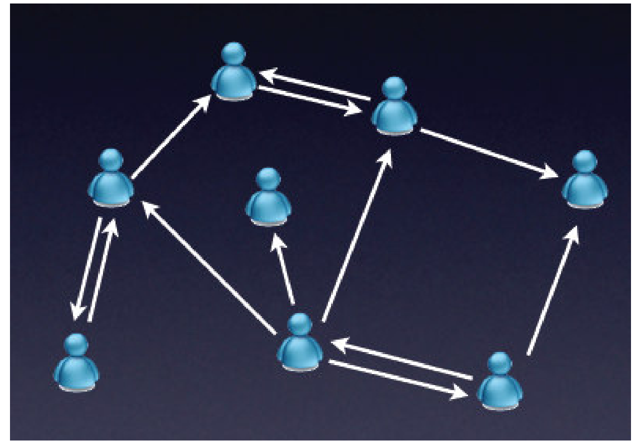


Part2.
4-step
approach

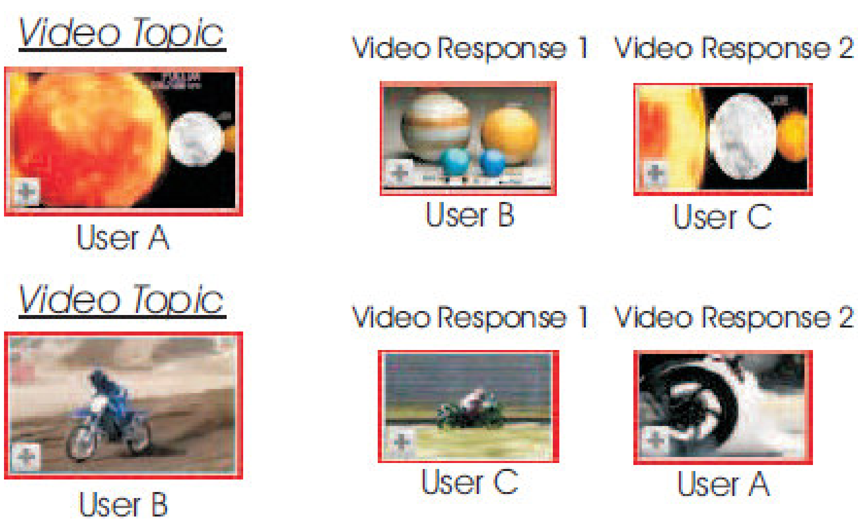
Part3.
Experimental
results

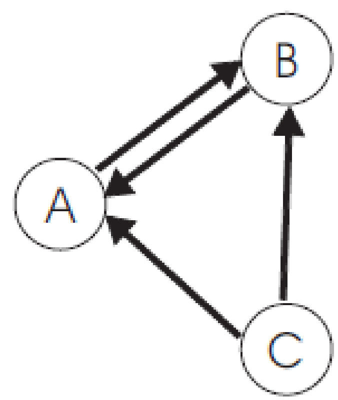
# Step1. Sampling video responses

- How people crawl social networks?
  - Pick known users
  - Crawl friends
  - Crawl new users found recursively



## Video response user graph



# Step1. Sampling video responses

- Crawls subject to rate-limiting
  - Use of a master-slave crawler with 10 client machines
- Effective performed a BFS of our graph
  - Seeds: list of top-100 most responded videos of all time
  - Follows links in both directions
  - Collect entire weakly connected components (WCCs)
- Collected 701,950 video responses and 381,616 video topics, 264,460 users in 7 days in January, 2008

# Step2. Create Test Collection

## Desired Properties

1) Have a significant number of users in each class

2) Include spammers and promoters which are aggressive in their strategies

3) Include a large number of legitimate users with different behavioral profiles

# Step2. Create Test Collection

- Users selected according to three strategies

  1) Manually identified 150 suspect in the top 100 most responded lists

  2) Randomly select 300 users from those who posted video responses to videos in the top 100 most responded lists

  3) Collected 400 users across 4 different levels of interaction

     - sent and received video responses

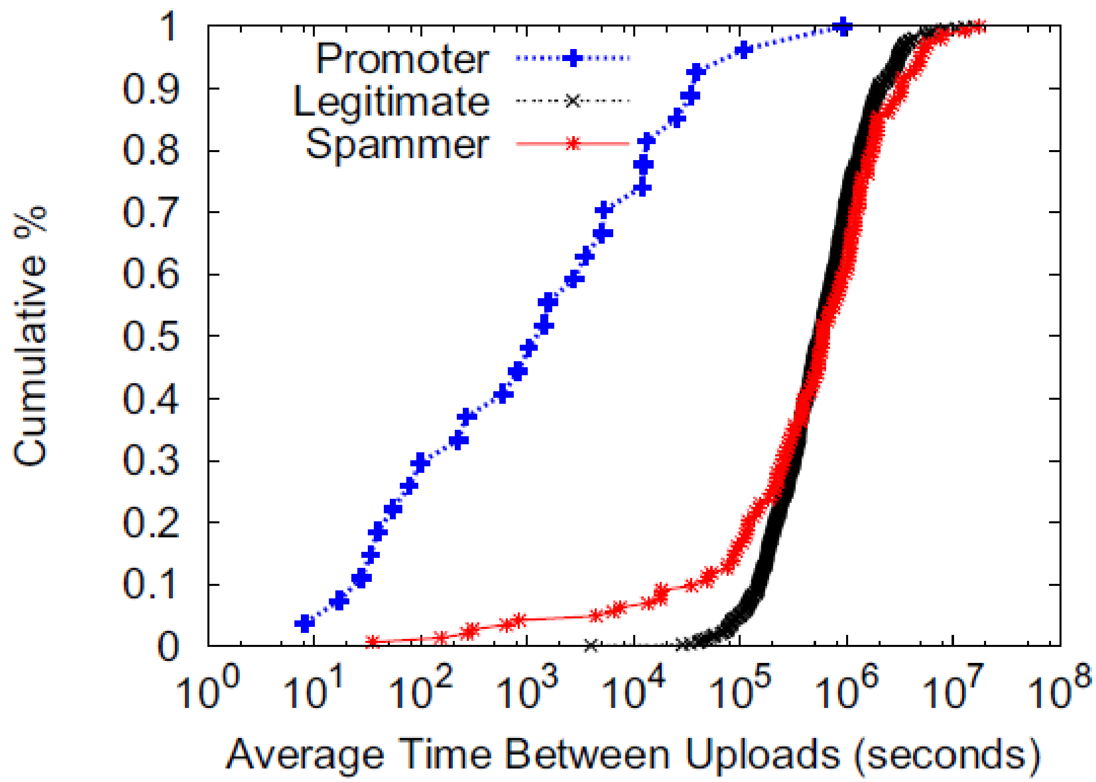
- Volunteers analyze users and videos
  - Conservative approach -> favor legitimate
  - Agreement in 97% of the analyzed videos

In total 829 users: 641 legitimate, 157 spammers, 31 promoters

# Step3. Attributes

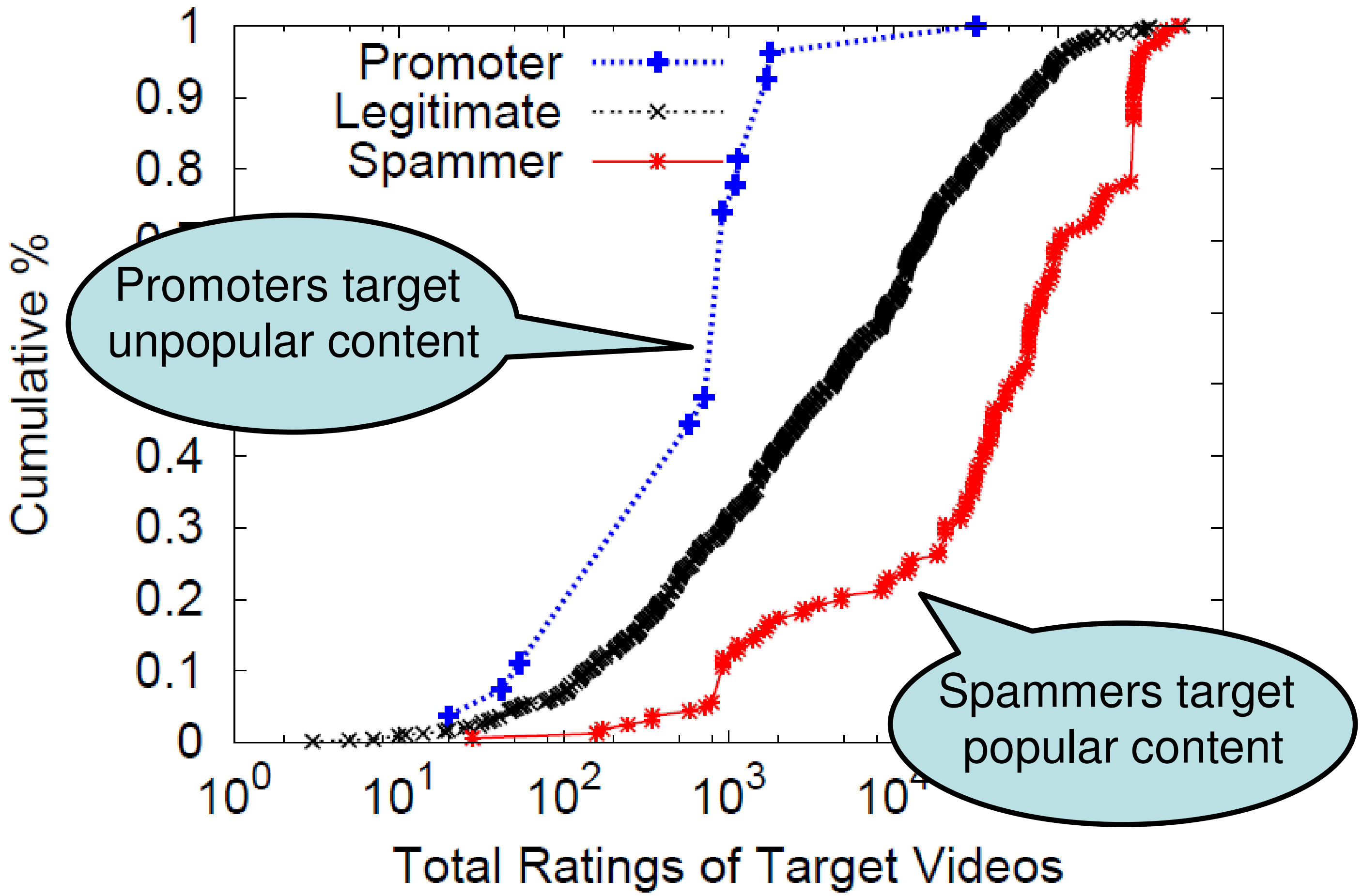
- User-Based:
  - number of friends, subscriptions, subscribers, favorites, videos watched, etc
- Video-Based:
  - duration, numbers of views received, comments, ratings, favorite marked, honors, external links, etc
  - 3 sets of videos: video topics, video responses, and all the videos
- Social Network:
  - clustering coefficient, betweenness, reciprocity, assortativity, UserRank (pagerank), etc

# Distinguishing classes of users (1)

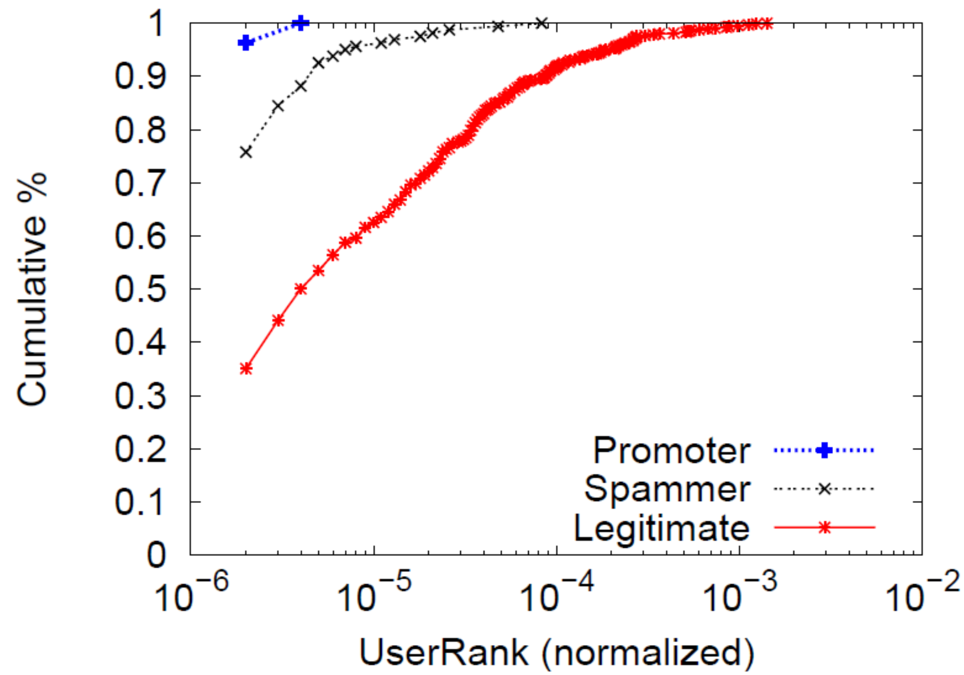



Promoters usually post several videos in a short period of time

# Distinguishing classes of users (2)



# Distinguishing classes of users (3)


Social network metrics have potential to separate classes apart

# Step3. Attributes

## Feature Selection: $\chi^2$ ranking

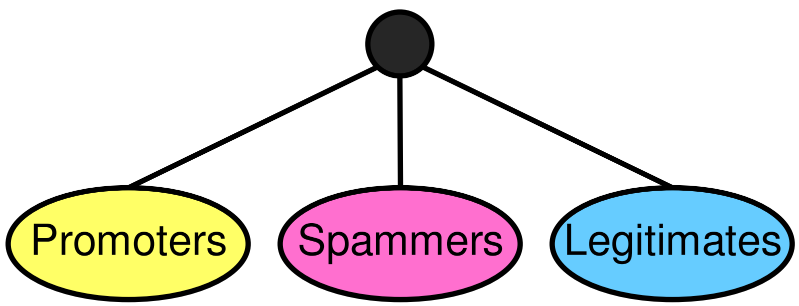
| Attribute Set | Top 10 | Top 20 | Top 30 | Top 40 | Top 50 |
|---|---|---|---|---|---|
| Video | 9 | 18 | 25 | 30 | 36 |
| User | 1 | 2 | 4 | 7 | 9 |
| SN | 0 | 0 | 1 | 3 | 5 |

Even low-ranked features have potential
to separate classes apart

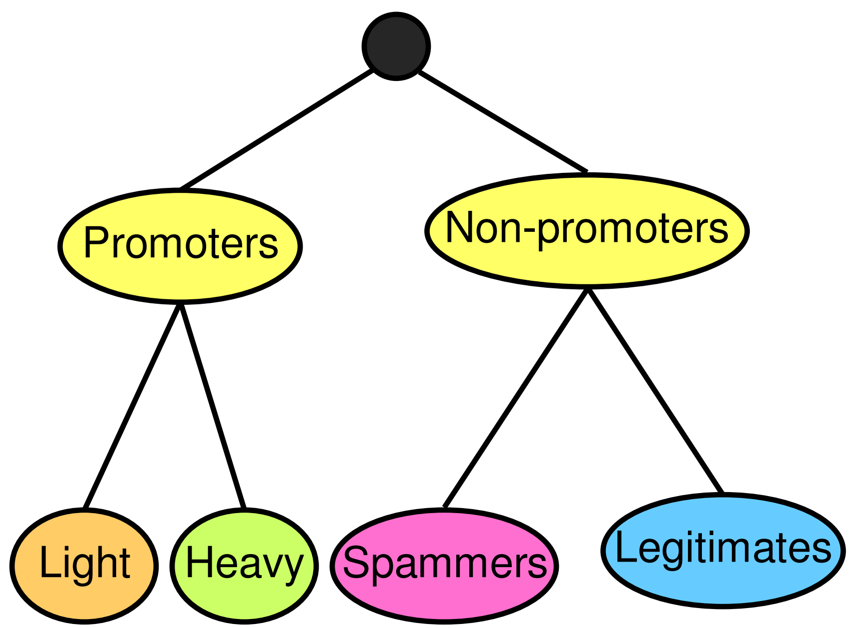
# Step4. Classification Approach

- SVM (Support vector machine) as classifier
  - Use all attributes
  - Two classification approaches

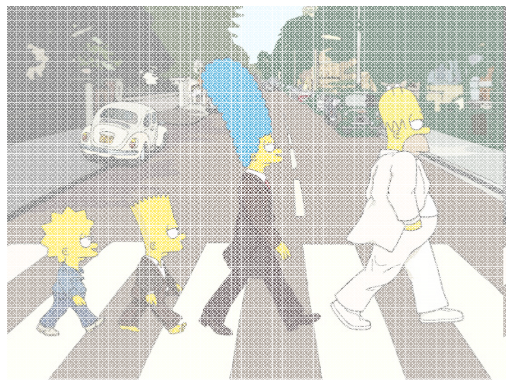
Flat




Hierarchical



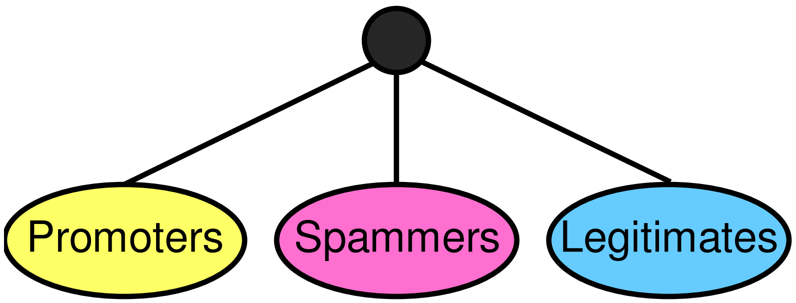
**Part1.**
**Motivation**
**& Problem**



**Part2.**
**4-step**
**approach**

**Part3.**
**Experimental**
**results**

# Flat Classification



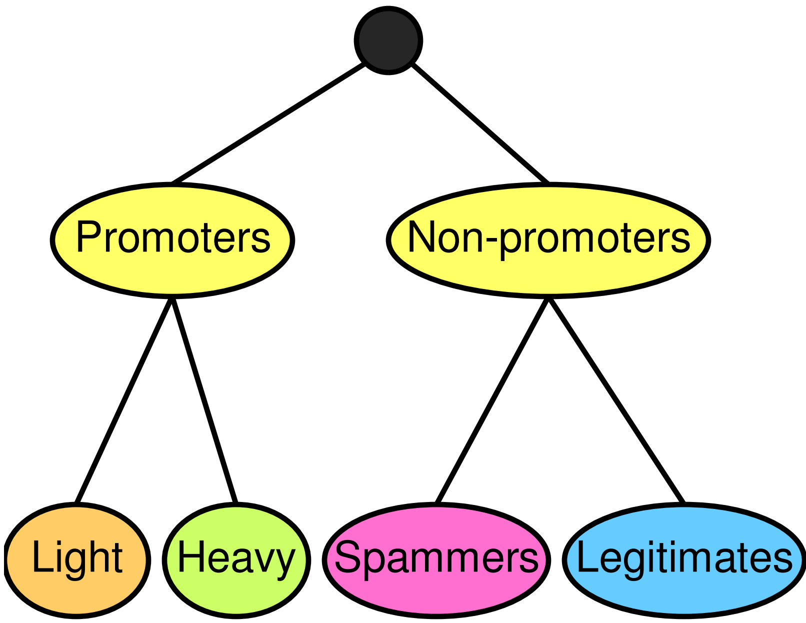

- Correctly identify majority of promoters, misclassifying few legitimate users.

- Detect a significant fraction of spammers but they are harder to distinguish from legitimate users
  - Dual behavior of some spammers

| | | Predicted | | |
|---|---|---|---|---|
| | | Promoter | Spammer | Legitimate |
| True | Promoter | **96.13%** | 3.87% | 0.00% |
| | Spammer | 1.40% | **56.69%** | 41.91% |
| | Legitimate | 0.31% | 5.02% | **94.66%** |

- Micro F1 = 88% (predict the correct class 88% of cases)

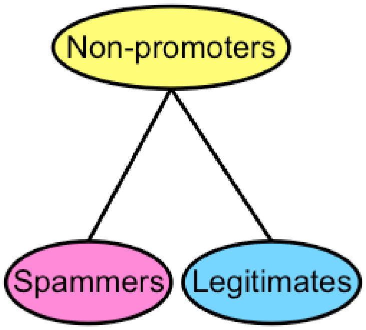
# Hierarchical Classification



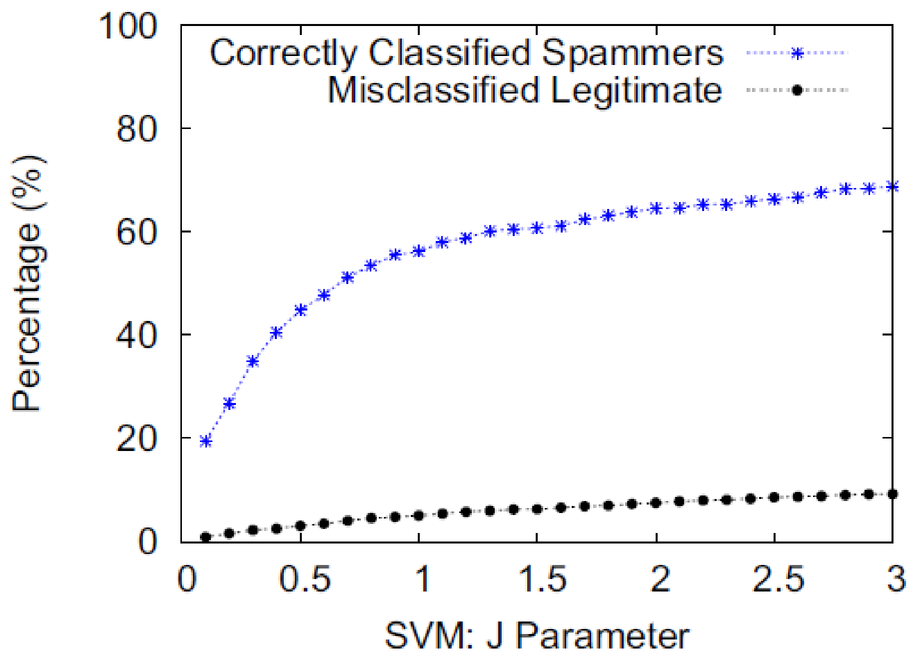

- Goal: provide flexibility in classification accuracy
- First Level:
  - Most promoters are correctly classified
  - Statistically indistinguishable compared with flat strategy

| | | Predicted | |
|---|---|---|---|
| | | Promoter | Non-Promoter |
| True | Promoter | **92.26%** | 7.74% |
| | Non-Promoter | 0.55% | **99.45%** |

# Distinguishing Spammers from Legitimate users



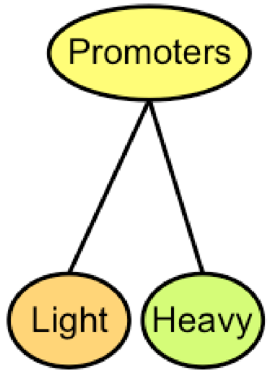

| | | Predicted | |
|---|---|---|---|
| | | Legitimate | Spammer |
| | Legitimate | **95.09%** | 4.91% |
| True | Spammer | 41.27% | **58.73%** |



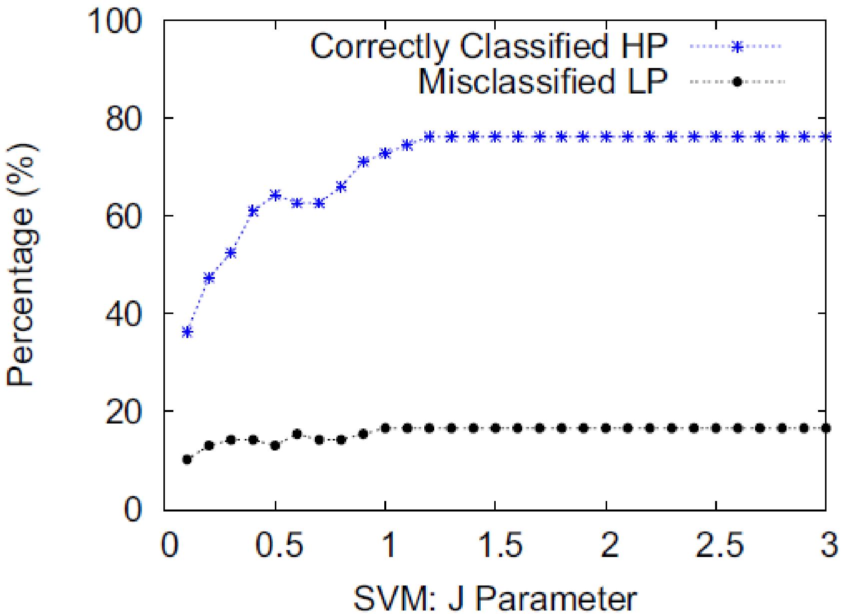

- **J = 0.1:** correctly classify 24% spammers, misclassifying <1% legitimate users
- **J = 3:** correctly classify 71% spammers, paying the cost of misclassifying 9% legitimate users

# Distinguishing Promoters

- **Heavy promoters** could reach the top-100 in one day
- **Light promoters** associated with a collusion attack



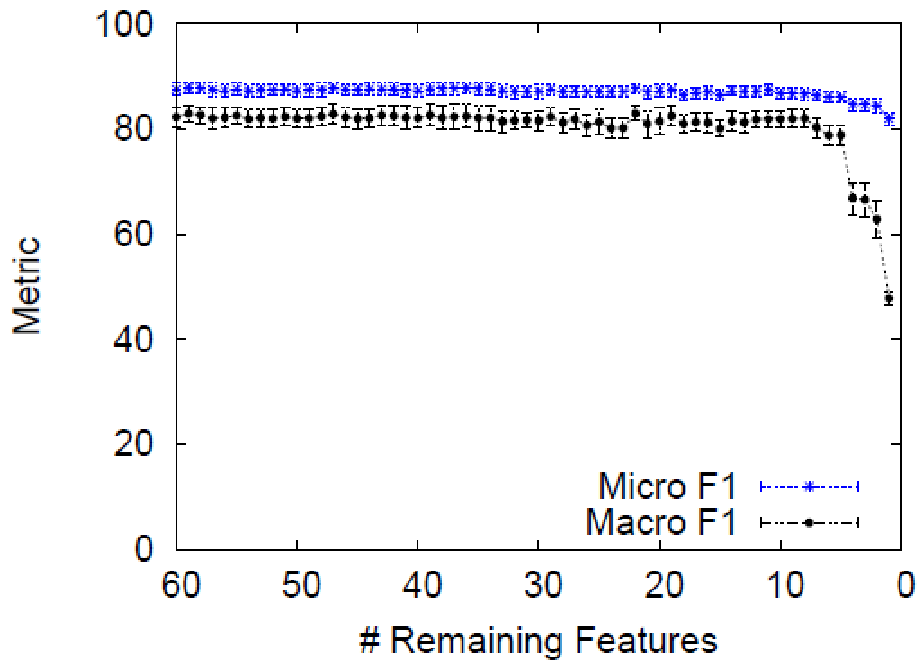

| | | Predicted | |
|---|---|---|---|
| | | Light Promoter | Heavy Promoter |
| True | Light Promoter | **83.33%** | 16.67% |
| | Heavy Promoter | 27.12% | **72.88%** |



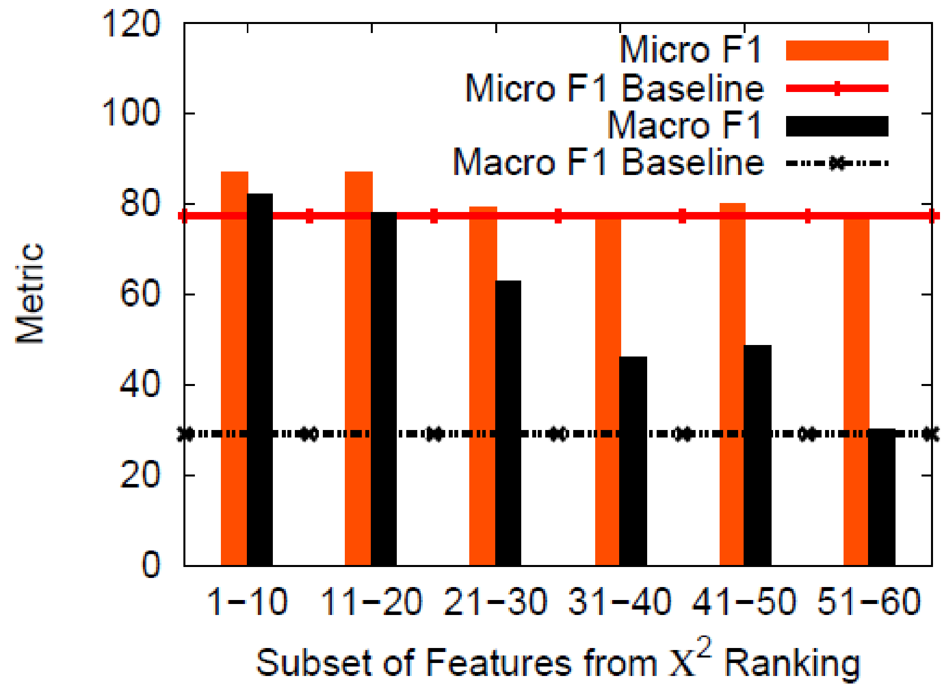

- **J = 0.1**: correctly classify 36% of heavy promoters at the cost of misclassifying 10% of light promoters
- **J = 1.2:** correctly classify 76% of heavy promoters at the cost of misclassifying 17% light ones

# Reducing the Attribute Set

## Scenario 1




## Scenario 2




Classification approach is effective even with a smaller, less expensive set of attributes

Different subsets of features can obtain competitive results

# Conclusions

- First approach to detect spammers and promoters
  - Attribute identification
  - Creation of a test collection
    - Publicly available at www.dcc.ufmg.br/~fabricio
  - Classification approach
    - Correctly identify majority of promoters
    - Spammers showed to be much harder to distinguish
      - trade-off between detect more spammers at the cost of misclassifying more legitimate users

# Discussion and Future Directions

- Other approaches that could be combined with ours
  - User Filtering
  - IP-Blocking, SMS account authentication
  - User reputation

- Future work:
  - Compare different classifiers and possibly combine them
  - Label users is expensive and time consuming
    - Evaluate semi-supervised classification methods

# Some Recent Publications

- Fabrício Benevenuto, Tiago Rodrigues, Virgílio Almeida, Jussara Almeida, Keith Ross. **Video Interactions in Online Video Social Networks**. ACM Transactions on Multimedia Computing, Communications and Applications (ACM TOMCCAP), 2009.
- Fabrício Benevenuto, Tiago Rodrigues, Meeyoung Cha, Virgílio Almeida. **Characterizing User Behavior in Online Social Networks**. ACM SIGCOMM Internet Measurement Conference (IMC'09), 2009.
- Fabrício Benevenuto, Tiago Rodrigues, Virgílio Almeida, Jussara Almeida and Marcos Gonçalves. **Detecting Spammers and Content Promoters in Online Video Social Networks**. In ACM SIGIR 2009.
- Fabrício Benevenuto, Fernando Duarte, Tiago Rodrigues, Virgílio Almeida, Jussara Almeida, Keith Ross. **Understanding Video Interactions in YouTube**. ACM Multimedia (MM'08), 2008.

# Questions?


fabricio@dcc.ufmg.br

http://www.dcc.ufmg.br/~fabricio