



## Detecting Tag Spam

Nicolas Neubauer

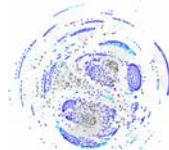
### Introduction

#### Social Tagging

Social tagging sites like del.iciou.us let users store their bookmarks online, along with descriptive keywords: „tags“



Sample tag network



Real tag network



Cooccurring tags



Tagging by spammer

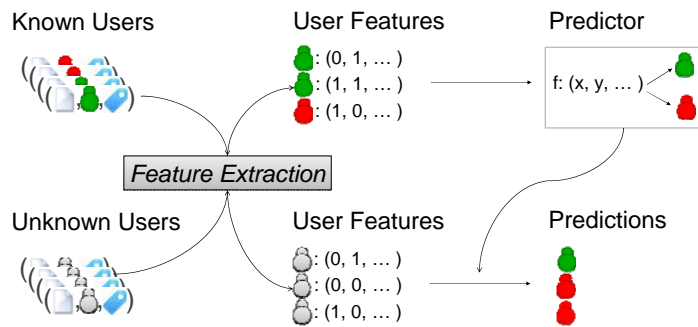
#### Tag Spam

The success of tagging has created incentives for spammers

### Question: How can we distinguish spammers from real users?

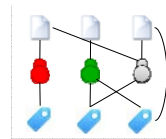
#### General Approach

Learn a model of spam and normal users so we can classify them in the future

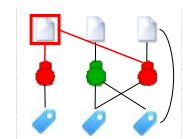


#### Feature Set I: Cooccurrences

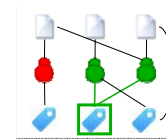
„Users which share resources or tags are likely to be similar in spaminess“



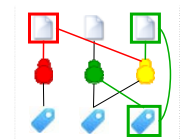
How do we use this information?



a) Propagate through documents



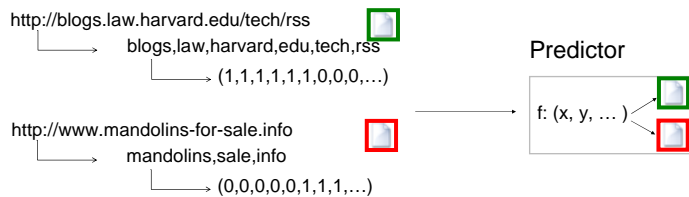
b) Propagate through tags



c) Iterate propagation

#### Feature Set II: URL Components

„Spam URLs tend to be composed of similar keywords“

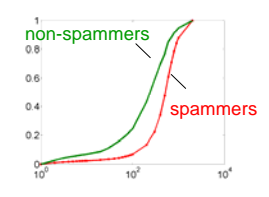


#### Feature Set III: Network Structure

„Differences in tagging behaviours lead to differences in graph structure“



Typical Structures in document/user networks (per tag) for real and spam tags

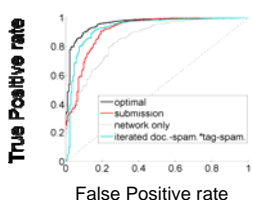


Quantifying network features: Avg of # connected components

### Results

#### Numerical Results

Competition: Classify 7000 users, based on 32000 known users (92% spammers)



Area under Curve (Performance)

- 0.96 optimal features
- 0.93 only features I (no content analysis required)
- 0.85 only feature set III (no labels required for feature extraction)

#### Interpretation

- + : Three different, successful approaches
- : Feature selection did not generalize well

Future Research:

- Feature Set I is an instance of Semisupervised Learning. How can advanced concepts from that field be transferred?
- Feature Set III is not as exact as the other features, but features can be created without previous labelling. How can network structure be further exploited for prediction?
- Spam classification is only one possible classification. Evaluate approach on other classification tasks.