Spam Detection by Extending and Propagating Trust and AntiTrust Scores

Tony Abou-Assaleh
&
Tapajyoti Das

GenieKnows.com

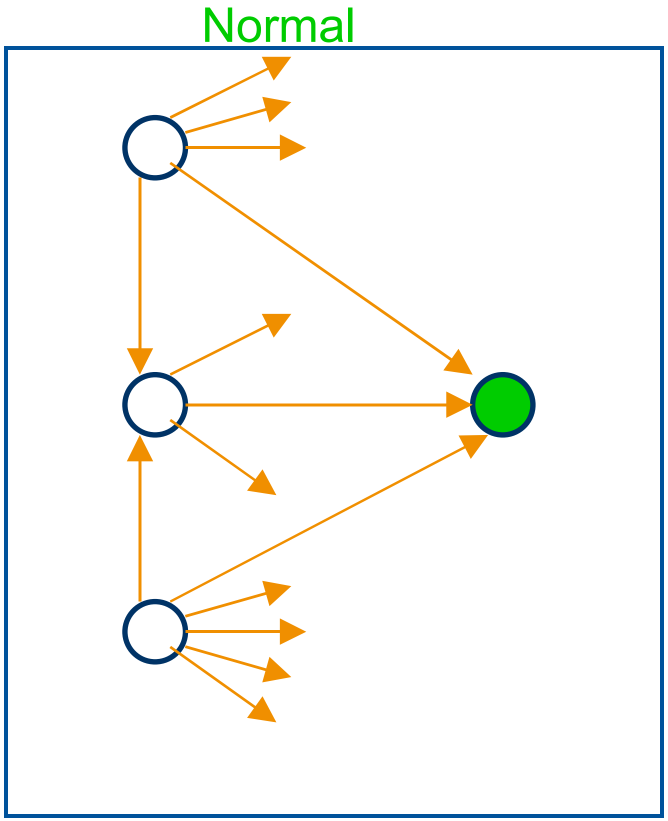
Web Spam Detection : Challenges?

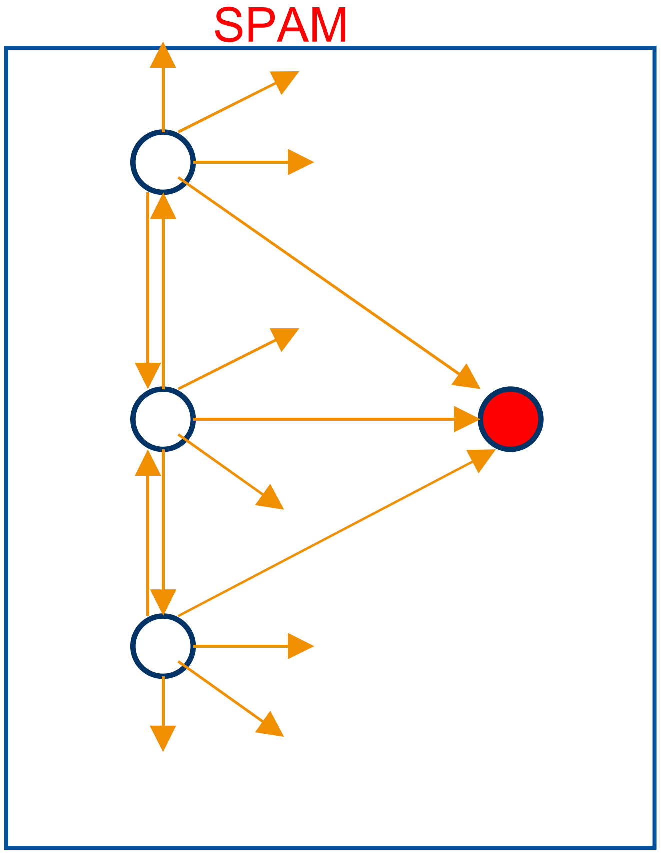
- **NO** objective definition
- Can be subtle and highly subjective
- Some spamming techniques are not visible/obvious.

Our Approach

- Based on Approximate Isolation of good pages
- Experiments using Hostgraph of the WEBSPAM-UK2006 dataset
- 674 manually labeled spam domains and 3100 good domains (.ac.uk, .gov.uk, .nhs.uk…)

Spam Detection : Variance of out-degrees of in-neighbours





Spam Detection

- Overlap of in-links and out-links.

  - Threshold of 5

- Extend the manually labeled spam set by adding all their in-neighbors.

- Extend the good core set by adding all out-neighbors.

Score Propagation

- Each domain was assigned three scores: **good**, **bad** and **combined**. Initialize score of good domains to +1, spam domains to -1, and zero for the rest.

- **Good Score** for a domain was the discounted average score of the in-neighbors, whereas for **Bad Score** it was the out-neighbors.

- The discount factor is α^{i}, where *i* is the iteration number and $\alpha = 0.20$

Combining the Scores

- The Combined Score was computed as :
  β*bad_score + (1-β)*good_score
  where β = 0.95;
- Domains with Negative Combined score were assigned Spam, others were assigned normal.
- Scores were scaled and shifted to fall within [0,1]

Results

- We have presented two results for evaluation :
  - Spam set comprising only variance based spam
  - Spam set including the manually labeled spam + automatic detected spam (i.e. variance based + overlap of links)
- Our approach achieves a F1 score of 0.92 with variance based method and 0.94 including the manual labels.
- Using the variance based method, we label 2787 domains as spam, while for the other method we label 3740 domains as spam.

Discussion

- **NO** Content based spam detection techniques. The manual labels serve as the content based spam.
- Combining both Trust and AntiTrust into the calculation of the Final score.

Search Engine Spam : Reality

Questions / Comments

Thank you