

# Spam and Popularity Ratings for Combating Link Spam

Mukesh Dalal

1533 Rio Grande St., Davis, CA 95616 USA

Mukesh.Dalal@gmail.com

## ABSTRACT

We present a new approach for propagating spam scores in web graphs, in order to combat link spam. The resulting spam rating is then used for propagating popularity scores like PageRank. Both propagations work even in presence of censure links that represent distrust. Initial testing using a C++ prototype on small examples show more reasonable results than other published approaches.

## Categories and Subject Descriptors:

H.3.3 [Information Search and Retrieval]: Search process.

**General Terms:** Algorithms, Management, Design.

**Keywords:** Search, Link spam, Trust, Distrust, PageRank, Web, WWW.

## 1. INTRODUCTION

Web spamming are dishonest practices that mislead search and indexing programs into giving undeserved search result rankings to web pages. We focus on link spamming [2] which takes advantage of link-based ranking algorithms such as PageRank [1] that gives a higher ranking to pages that are linked from other highly ranked pages. An early approach to combat link spam introduced TrustRank [3] that, like PageRank, flows out through links from a set of manually-identified trusted sites. In addition to trust, [5] also propagates distrust backward through links from a manually-identified set of spam pages.

Although trust and distrust propagation have also been studied in the more general context of reputation systems [4], these notions are associated there with links, such as “A trusts B” and “B distrusts C”. To avoid confusion with this more common usage, we use spam and popularity scores for web pages, instead of trust and mistrust scores. Thus, TrustRank and PageRank are examples of popularity score, while the distrust score of [5] is an example of spam score. Just like in reputation systems, we also allow an optional trust value to be associated with each link.

We make four contributions in this paper. First, we propose a new approach for propagating spam scores that seems to improve over other published approaches. Second, we leverage spam rating in propagating popularity scores. Third, we use optional trust value of links in both spam and popularity propagation. Fourth, we allow both positive and negative scores for spam, popularity and trust.

We represent the link structure of the web (web graph) by the adjacency matrix  $M$ , where  $M[a, b]$  is the sum of trusts of all links from page  $a$  to page  $b$ . In the absence of trust scores,  $M[a, b]$  is 1 if there is a link from  $a$  to  $b$ , and 0 otherwise.

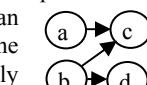
Copyright is held by the author/owner(s).

WWW 2007, May 8–12, 2007, Banff, Alberta, Canada.

ACM 978-1-59593-654-7/07/0005.

## 2. PROPAGATING SPAM SCORES

The basic idea is that a page linking to other pages with high spam scores should also get a high spam score. Thus, spam scores should be propagated backward through links. This basic idea need to be refined in the following ways:

1. *Row normalization:* A page should not get a higher spam score just because it links to more pages. Thus, the total trust of all outgoing links should be normalized across all pages to either 1 (if there are outgoing links) or 0 (otherwise).
2. *Column normalization:* Since spam score of a page flows backwards through all incoming edges, the total trust of all incoming links should be normalized across all pages (again, either 1 or 0). The matrix obtained from first row and then column normalization of  $M$  is called backward propagation matrix  $B$ . For example, the top-right quadrant  $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$  of  $M$  for the adjoining web graph gets transformed to  $\begin{bmatrix} 2/3 & 0 \\ 1/3 & 1 \end{bmatrix}$  in  $B$ ; all other entries in  $M$  and  $B$  are 0. Since   $B$  determines the backward propagation, the spam score of Page  $c$  will flow double more to Page  $a$  than to Page  $b$ . We now argue that this is the most reasonable behavior. Suppose only Page  $c$  has been identified as a spam page, that is, with a high spam score. Now Page  $a$  links only to a spam page, while Page  $b$  links to a spam page and a possibly non-spam page. Thus, Page  $a$  should get a higher spam score than Page  $b$ , consistent with our approach. Note that none of the other published approaches seem to provide this intuitively reasonable flow.
3. *Decay and bias:* Just like the use of decay and bias (for both conceptual and mathematical reasons like convergence) in computing PageRank, propagation of spam scores uses spam decay factor  $\beta$  and spam bias (column) vector  $v$ . However, in contrast to the PageRank bias, a uniform spam bias is not at all useful. A non-uniform spam bias could be obtained from an a priori spam scoring of selected web pages, either manually or by using some automated approach, say, based on content analysis. We have obtained reasonable results for  $\beta$  between 0.3 and 0.9.

The spam rating  $s$  (a column vector of spam scores) based on this backward propagation is obtained by solving the system of equalities  $(I - \beta B)s = v$ , where  $I$  is the  $N \times N$  identity matrix and  $N$  is the number of pages. Since  $(I - \beta B)$  is a strictly-diagonally dominant matrix, the solution exists and can be calculated efficiently using PageRank style optimizations. The vector  $s$  is often rescaled so that the maximum value is 1.

### 3. PROPAGATING POPULARITY

Since a page linked from other pages with high popularity scores should get a high popularity score, popularity scores should be propagated forward through links. We augment this standard approach by using the spam score of a page to somewhat repel the flow of popularity to it. In general, a page with higher spam score will repel more, effectively sending that popularity to pages with lower spam scores. In particular, the flow of popularity is modified in two ways:

1. Each entry in  $M[a, b]$  in the adjacency matrix is multiplied by  $e^{-s(b)}$ , where  $s(b)$  is the scaled spam score of Page b. This transformation serves two purposes: trust in links to nodes with higher spam score is reduced more and negative spam scores are automatically handled. The resulting matrix is row-normalized, as usual, to obtain the forward adjacency matrix F.
2. Each entry  $u(a)$  of the popularity bias (column) vector u is also multiplied by  $e^{-s(b)}$ , for the same reasons as above.

The popularity rating p (a column vector of popularity scores) based on this forward propagation is obtained by solving the system of equalities  $(I - \alpha F^T)p = u * e^{-s}$ , where  $\alpha$  is the popularity decay factor. Since  $(I - \alpha F^T)$  is also a strictly-diagonally dominant matrix, the solution exists and can be calculated efficiently using PageRank style optimizations. We have obtained reasonable results with  $\alpha$  between 0.3 and 0.9 (just like  $\beta$ ).

### 4. CENSURE LINKS & NEGATIVE BIAS

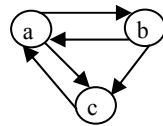
Typical propagation approaches treat links as endorsement, that is, a link from page a to b is considered an endorsement of page b by a, except for one special case: links with attribute rel="nofollow" are ignored by many search engines, as if those links do not exist. However, there is currently no way to create links that censure other pages, that is, provide negative endorsement. Such censure links would allow, for example, page a to link to page b identifying b as a spam site. We allow censure links by using negative trust values; the positive trust values represent regular endorsement links, while trust value of 0 represents "nofollow" links. While such negative trust links expressing distrust are often used in the more general reputation systems, they have not been used in spam or popularity propagation through web graphs.

We also allow negative values in both spam and popularity bias vectors. A negative spam bias represents a priori (possibly manual) determination of a quality node, while the interpretation of a negative popularity bias is not yet clear.

Interestingly, the spam and popularity propagation approaches described above continues to work for censure links and negative biases. Intuitively, a censure link to page a is treated as an endorsement link to a hypothetical page whose score (spam and popularity) is negative of a's score. However, we make a minor modification to allow optional discounting of censure links in propagating popularity. In particular, each negative value in adjacency matrix M is first multiplied by the negative discount factor  $\delta$  (between 0 and 1) in obtaining the forward propagation matrix F. This discounting prevents several popular pages to group together to maliciously demote another page.

### 5. A TOY EXAMPLE

We now present a complete example to illustrate the computation of spam and popularity ratings. Consider the adjoining web graph



$$\text{with adjacency matrix } M = \begin{bmatrix} 0 & 1 & 0.5 \\ 1 & 0 & -0.8 \\ 1 & 0 & 0 \end{bmatrix},$$

spam bias vector  $v = [1 \ 0 \ 0]^T$ , popularity bias

vector  $u = [1 \ 1 \ 1]^T$ , spam decay factor  $\beta=0.3$ , popularity decay factor  $\alpha=0.85$ , and negative discount factor  $\delta=0.5$ . Note that the link from page b to c is a censure link and that page a has been a priori identified as a spam node.

$$\text{The backward propagation matrix } B \text{ then is } \begin{bmatrix} 0 & 1 & .429 \\ .357 & 0 & -.571 \\ .643 & 0 & 0 \end{bmatrix}$$

and the rescaled spam rating  $s = [1 \ 0.074 \ 0.193]^T$ . Note that b has a low spam score partly due to the censure link. The

$$\text{forward propagation matrix } F \text{ is } \begin{bmatrix} 0 & .64 & .36 \\ .849 & 0 & -.151 \\ 1 & 0 & 0 \end{bmatrix} \text{ and the}$$

$$\text{rescaled popularity rating } p = [0.864 \ 1 \ 0.26]^T.$$

### 6. CONCLUSIONS

We have developed a C++ prototype for propagating spam and popularity scores, and have tested it on several published examples with reasonable results. However, we still need to test and validate this approach on large realistic web graphs.

Using negative trust scores may invite retaliation from censured pages. Thus, it is important to develop approaches for providing anonymous trust scores. Also, web pages should be encouraged to provide trust scores for links, probably by incorporating "percentage rated" in spam or popularity scores.

### 7. REFERENCES

- [1] Brin, S. and Page, L. The Anatomy of a Large-Scale Hypertextual Web Search Engine. WWW7 / Computer Networks, 30 (1-7). 107-117, 1998.
- [2] Gyongyi, Z. and Garcia-Molina, H. Web Spam Taxonomy. First International Workshop on Adversarial Information Retrieval on the Web, 2005.
- [3] Gyongyi, Z., Garcia-Molina, H. and Pedersen, J. Combating Web Spam with Trustrank. Proceedings of the 30th International Conference on Very Large Data Bases (VLDB), 2004.
- [4] Jøsang, A., Ismail, R. and Boyd, C. A Survey of Trust and Reputation Systems for Online Service Provision. Decision Support Systems, 2005.
- [5] Wu, B., Goel, V. and Davison, B.D., Propagating Trust and Distrust to Demote Web Spam. Workshop on Models of Trust for the Web, Edinburgh, Scotland, 2006.