Winnowing Wheat from the Chaff: Propagating Trust to Sift Spam from the Web

Lan Nie Baoning Wu Brian D. Davison
Department of Computer Science & Engineering
Lehigh University
Bethlehem, PA 18015 USA
{lan2,baw4,davison}@cse.lehigh.edu

ABSTRACT

The Web today includes many pages intended to deceive search engines, and attain an unwarranted result ranking. Since the links among web pages are used to calculate authority, ranking systems would benefit from knowing which pages contain content to be trusted and which do not. We propose and compare various trust propagation methods to estimate the trustworthiness of each page. We find that a non-trust-preserving propagation method is able to achieve close to a fifty percent improvement over TrustRank in separating spam from non-spam pages.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing

General Terms

Algorithms, Performance

Keywords

Web search engine, authority, trust, spam

1. INTRODUCTION

In the early days of the Web, a web search engine could be perfectly objective, examining only the content of a page and returning those pages whose content best matched the query. However, the growth of the Web meant that thousands or millions of pages were often considered relevant, leading to the use of links as votes or recommendations to estimate the authority of web pages.

With some knowledge of how search engines function, it is possible to manipulate the results of a search engine query by adding keywords to the content or by creating links from other pages to the target page.The use of such techniques, called search engine spam, can lead to inappropriately high rankings for the target pages while degrading overall query results.

Traditional link analysis techniques consider the content and links on all pages. However, given the adversarial nature of today's web, it would be advantageous to know which pages are trustworthy, so that they may be promoted in authority calculations [5].

Gyöngyi et al.'s TrustRank [4] was one of the first mechanisms to calculate a measure of trust for Web pages. It is based on the idea that good sites seldom point to spam sites and people trust these good sites. It uses a human-selected seed set of highly trustworthy nodes, and then calculates a personalized PageRank [1] in which all jump probability is distributed only to the seed set:

$$TR(i) = d \sum_{j:j\to i} \frac{TR(j)}{O(j)} + \begin{cases} (1-d)\frac{1}{|\tau|} & \text{if } i \in \tau \\ 0 & \text{if } i \notin \tau \end{cases} \quad (1)$$

where $TR(i)$ is the TrustRank score for page i and τ is the seed set. $TR(i)$ will be initialized to $\frac{1}{|\tau|}$ if $i \in \tau$ and 0 otherwise. Gyöngyi et al. iterate 20 times with d set to 0.85.

Those pages that are reachable via the directed graph from a seed node accumulate some trust; the better linked a page is to the seed set, the higher the trust score calculated. TrustRank promotes trustworthy pages, and demotes untrustworthy pages (e.g., spam pages). However, it is not clear that trust should flow in the same way as authority (as demonstrated by Guha et al. [3] in a person-to-person trust network). More recently, Wu et al. [6] proposed using different mechanisms to propagate trust among web pages. In addition, they incorporated distrust into the model using reverse propagation.

In this poster we demonstrate 1) the creation and use of a novel trust evaluation metric that incorporates spam and non-spam page measures; and, 2) the comparison of propagation mechanisms for trust and distrust, and their combination in a unified model.

2. PROPAGATION OF TRUST

TrustRank propagates trust identically to how PageRank propagates authority. While the use of a known seed set is valuable, we consider here some alternatives to the usual method of propagation. The first is, for each parent, how to divide its score amongst its children ("splitting"). The other is, for each child, how to calculate the overall score given the shares from all of its parents ("accumulation"). In the case of TrustRank, a parent's trust score is equally distributed among its children, and a child's overall trust score is the sum of the shares from all of its parents.

In particular, with respect to trust splitting, we question the need to give less weight to recommendations made by one entity simply because the entity made more recommendations. One straightforward alternative is to grant each child the full measure of trust assigned to the parent rather than equally splitting. Thus, we consider the two choices:

- **Equal Splitting (Eq):** a node i with $O(i)$ outgoing links and trust score $Trust(i)$ will give $\frac{Trust(i)}{O(i)}$ to each child.
- **Constant Splitting (Con):** a node i with trust score $Trust(i)$ will give $Trust(i)$ to each child;

Additionally, in either case, a child's trust need not be simply the sum of the parents' trust. An alternative is to use the maximal trust sent by any parent. We investigate both choices for accumulation: **Simple Summation (Sum)** in which we sum the trust values

SIGIR'07, July 23–27, 2007, Amsterdam, The Netherlands.
ACM 978-1-59593-597-7/07/0007.

Distrust Algorithm	Trust Algorithm			
	Con_ Sum	Eq_ Sum	Con_ Max	Eq_ Max
Con_Sum	4.00	2.84	-0.54	0.12
Eq_Sum	4.12	2.94	-0.38	0.26
Con_Max	4.00	2.90	-0.34	0.12
Eq_Max	**4.13**	2.93	-0.34	0.27

Table 1: Average increase in bucket gap between normal and spam pages for each combination of trust and distrust methods.

Distrust Algor.	Trust Algorithm							
	Con_Sum		Eq_Sum		Con_Max		Eq_Max	
	No	Sp	No	Sp	No	Sp	No	Sp
Con_Sum	23.3	-5.7	-2.0	-5.7	-110	-5.7	-98.8	-4.9
Eq_Sum	23.4	-5.7	-1.9	-5.7	-110	-5.7	-98.9	-5.0
Con_Max	23.3	-5.7	-2.4	-5.7	-110	-5.7	-98.8	-4.9
Eq_Max	**23.4**	**-5.7**	-2.4	-5.7	-110	-5.7	-98.8	-5.0

Table 2: Increase in the number of spam (Sp) and normal (No) pages within the top ten buckets.

from each parent, and **Maximum Share (Max)** in which we use the maximum trust value sent by a parent. Each of these propagation policies is applicable to trust and distrust, except that distrust propagates along the reverse web graph from spam seed sets. By using the above choices, the equation for calculating trust (or distrust) will incorporate modifications to Equation 1. For example, if using "Constant Splitting" and "Simple Summation" for trust propagation (denoted as Con_Sum), the equation will become:

$$\mathit{Trust}(i) = d \sum_{j:j\to i} \mathit{Trust}(j) + \begin{cases} (1-d)\frac{1}{|\tau|} & \text{if } i \in \tau \\ 0 & \text{if } i \notin \tau \end{cases} \quad (2)$$

The two propagation processes result in two scores associated with each page. An overall trust score can be generated by subtracting the distrust score from the trust score, in the form of $\mathit{Total}(i) = \mathit{Trust}(i) - \alpha \times \mathit{Distrust}(i)$ where $\mathit{Total}(i)$ represents the overall trustworthiness for page i, $\mathit{Distrust}(i)$ is the calculated distrust for page i, and α is a weighting factor.

3. EVALUATION OF APPROACH

For our experiments we use the UK-2006 dataset [7]—a recent crawl of the .uk top-level domain containing 77M pages from 11,392 different hosts. Host labels are also available [2], in which 767 hosts are marked as spam, 7,472 as normal and 176 hosts as undecided. The remaining 2,977 hosts are unlabeled.

We use PageRank (actually HostRank since it is calculated within the host graph) as our baseline ranker. Following TrustRank's approach, we generate the list of sites in decreasing order of their PageRank values and partition them into 20 buckets, with each bucket containing hosts whose PageRank values sum to 1/20th of the total. For each proposed ranking, we can also calculate a corresponding ranking list which is then divided into 20 buckets so that each has an identical number of elements as the corresponding PageRank bucket. The first 10 PageRank buckets hold just under 10% of all hosts.

Our goal is to simultaneously demote spam sites and boost normal sites. This means that when evaluating performance, we need to track the movements of both normal and spam sites: the further they move away from another, the better the overall performance. Therefore, we calculate the change in gap (measured in buckets) between the average positions of normal and spam pages in a new ranking versus the baseline PageRank. We perform a stratified tenfold cross-validation to estimate the performance of each method on one-tenth of the spam and good hosts.

For each possible combination of splitting and accumulation in both trust and distrust propagation, we found the best-performing weighting factor α and used it to report the change in gap between spam and good pages, shown in Table 1. We find that using "Constant Splitting" with "Simple Summation" for trust propagation and "Equal Splitting" with "Maximum Share" for distrust propagation will achieve the best performance, moving normal and spam pages 4.13 buckets further apart on average compared to PageRank. From this table, we can tell that using "Simple Summation" rather than "Maximum Share" for accumulating trust will greatly improve the performance; in addition, using "Constant splitting" propagate trust outperforms the default "Equal Splitting", which confirms our intuition on the influence that out-degree should have on the trust passed to a child. TrustRank, in comparison, achieves a gap increase of only 2.83 buckets.

Table 2 considers the change in composition of the top 10 buckets compared to PageRank. The optimal combination again uses "Constant Splitting" and "Simple Summation", placing 23.4 more normal hosts in the top 10 buckets while moving out 5.7 spam hosts (accounting for all spam present in the top buckets). TrustRank also removes the 5.7 spam hosts, but 2.2 more normal hosts are pushed out from the top buckets as well. In contrast, the worst approach will push 110 normal hosts and 5.7 spam hosts out of the top buckets, which means that although this approach demotes spam, it also hurts many normal hosts.

4. CONCLUSION

We have compared various trust propagation methods to estimate the trustworthiness of web pages. Utilizing a novel trust evaluation metric that incorporates spam and non-spam measures, we show that a non-trust-preserving propagation method can dramatically improve upon TrustRank.

Acknowledgments

This material is based upon work supported by Microsoft Live Labs ("Accelerating Search") and the National Science Foundation under CAREER award IIS-0545875. We also thank the Laboratory of Web Algorithmics, Università degli Studi di Milano and Yahoo! Research Barcelona for making the UK-2006 dataset and labels available.

5. REFERENCES

[1] S. Brin, R. Motwani, L. Page, and T. Winograd. What can you do with a web in your pocket? *Data Engineering Bulletin*, 21(2):37–47, 1998.

[2] C. Castillo, D. Donato, L. Becchetti, P. Boldi, M. Santini, and S. Vigna. A reference collection for web spam. *ACM SIGIR Forum*, 40(2), Dec. 2006.

[3] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In *Proceedings of the 13th International World Wide Web Conference*, pages 403–412, New York City, May 2004.

[4] Z. Gyöngyi, H. Garcia-Molina, and J. Pedersen. Combating web spam with TrustRank. In *Proceedings of the 30th International Conference on Very Large Data Bases (VLDB)*, Toronto, Canada, 2004.

[5] L. Nie, B. Wu and B. D. Davison. A Cautious Surfer for PageRank. In *Proceedings of the 16th International World Wide Web Conference (WWW)*, pages 1119–1120, Banff, Canada, May 2007.

[6] B. Wu, V. Goel, and B. D. Davison. Propagating trust and distrust to demote web spam. In *Proceedings of the WWW2006 Workshop on Models of Trust for the Web (MTW)*, Edinburgh, Scotland, May 2006.

[7] Yahoo! Research. Web collection UK-2006. http://research.yahoo.com/. Crawled by the Laboratory of Web Algorithmics, University of Milan, http://law.dsi.unimi.it/. URL retrieved Oct. 2006.