Using the ProT Nordic Web Dataset

Ola Ågren

Abstract

In this paper we present a free dataset, usable for testing web search engines. The dataset corresponds to a snapshot of the Nordic part of the Internet back in early 2007 and is highly abstracted, with numbers representing each web page. The released dataset consists of three parts; a graph, 76 sets of pages containing each tested word combination, and some files to use when calculating relevance of the resulting sets of algorithms/search engines. We also present statistics for some search engine algorithms.

Contents

1 Introduction **3**
- 1.1 The Dataset . . . 3
- 1.2 Outline . . . 3

2 Statistics **4**
- 2.1 Relevance . . . 4
- 2.2 Coverage . . . 4
- 2.3 Adjusted Relevance . . . 4
- 2.4 Reporting Relevance Values . . . 5
- 2.5 Stability Tests . . . 5

3 Files **6**
- 3.1 Link Files . . . 6
- 3.2 Original Page Sets . . . 6
- 3.3 Assessment Data . . . 7

4 Sample Results **8**
- 4.1 Straight Link Count . . . 8
- 4.2 PageRank . . . 8
- 4.3 Topic-Sensitive PageRank . . . 9
- 4.4 HITS . . . 10
- 4.5 ProT . . . 12
- 4.6 S^2ProT . . . 12

5 Discussion **16**

A Maximum Spread of Relevance **19**

List of Tables

1 Relevance values for some common web-search algorithms 15
2 Minimum and maximum possible relevance numbers 19

List of Figures

1 Minimum and maximum values of relevance, given a certain coverage. 5
2 Sample link file 6
3 Original page set example 6
4 Grade file 7
5 Spread file 7
6 PageRank Relevance stability 9
7 Topic-Sensitive PageRank Relevance stability 10
8 HITS Authority Relevance stability 11
9 HITS Hub Relevance stability 11
10 ProT Relevance stability 13
11 S^2ProT Relevance stability 13
12 Ranking order: Spearman Footrule Distance 14
13 Ranking order: Spearman ρ 14
14 Relevance, limits, confidence interval and coverage for each algorithm 15

1 Introduction

When the original ProT Web search algorithm was devised and tested back in 2005, we were unable to find a suitably large freely available dataset containing both a link list and a set of gradings for a number of keywords. The small dataset used in [Åg06] was created by letting a web spider mirror everything off the web server of the Department of Computing Science at Umeå University, and then grading the top 5 pages for each tested algorithm and keyword.

This was deemed much too small for further testing, leading to using a mirror of all web servers at Umeå University. The tests done showed that even this dataset was too small and, moreover, it had a distinct bias towards course and research based pages. The solution was to increase the number of servers considerably, leading to the dataset that was used in [Åg08] and is the main focus of this report.

1.1 The Dataset

The dataset was collected in January and February 2007. It consists of all web pages found by our web spider within the Nordic countries, i.e. Denmark, Finland, Iceland, Norway, and Sweden. This subset of the Internet was chosen because these countries have had access to the Internet for a long time, they contain a mix of old and new, academic and commercial, web servers, and provide fast access from our location as described in [Åg08].

This database contains 3,087,531 web addresses, of which 478,985 pages contain hyper-links that point to another server. It contains a total of 37,245,054 unique hyper-links, of which 3,889,216 are non-local. The corresponding unweighted non-local adjacency matrix has a dominant eigenvalue of $\lambda_1 \approx 49.135476$.

All in all, 8,054,200 stemmed words of at least four characters appear 221,259,520 times in 727,757 of the pages, leading to an average of just over 304 unique words per page in the set. Of these, 76 of the words have been assessed and are given in Section 3.2 on page 6 and Section 3.3 on page 7.

1.2 Outline

The general outline of this report is as follows: Section 2 describes how this dataset can be used for testing web page algorithms while Section 3 describes the files that are available, Section 4 contains some sample data for a number of algorithms, and some discussions as well as future works can be found in Section 5.

2 Statistics

2.1 Relevance

The dataset is based on 76 words, with between 1 and 20 assessments of the pages that was in the top 10 list for the three algorithms that was assessed in [Åg08]. Five different grades were used in the assessment, corresponding to unknown (ignored in all calculations), complete lack of relevance (i.e. 0), some relevance (i.e. 0.5), relevant (i.e. 0.8) and very relevant (i.e. 1). The corresponding grade counts are $\text{grade}_0 \ldots \text{grade}_4$.

Other algorithms will in general yield rankings that have at least some of the top elements in common with those in the dataset. These are used as a basis for calculating the average relevance ($\overline{r}$) for that algorithm as in Eq. 1. A page that has not been assessed is deemed to belong to grade_5 (with the same count as the total number of assessments made for that word) and must be handled separately, since they are used to show the lower and upper limits of the relevance of the algoritm (equations 2 and 3, respectively). A fast way of calculating the standard deviation (σ) of the relevance is given in Eq. 4.

$$\text{sum}_g = 0.5\text{grade}_2 + 0.8\text{grade}_3 + \text{grade}_4$$

$$\text{count}_g = \text{grade}_1 + \text{grade}_2 + \text{grade}_3 + \text{grade}_4$$

$$\overline{r} = \text{relevance} = \frac{\text{sum}_g}{\text{count}_g} \tag{1}$$

$$\overline{r}_{\min} = \text{relevance}_{\min} = \frac{\text{sum}_g}{\text{count}_g + \text{grade}_5} \tag{2}$$

$$\overline{r}_{\max} = \text{relevance}_{\max} = \frac{\text{sum}_g + \text{grade}_5}{\text{count}_g + \text{grade}_5} \tag{3}$$

$$\sigma = \sqrt{\frac{\overline{r}^2\text{grade}_1 + (0.5 - \overline{r})^2\text{grade}_2 + (0.8 - \overline{r})^2\text{grade}_3 + (1 - \overline{r})^2\text{grade}_4}{\text{count} - 1}} \tag{4}$$

2.2 Coverage

Coverage is the number of top ten pages by an algorithm that is found in the dataset, either given as a straight count (as used in the relevance adjustment, see Section 2.3) or as a percentage (the count divided by 760).

2.3 Adjusted Relevance

The relevance number calculated using Eq. 1 can be used directly, as long as the algorithms have more or less the same coverage. There is, however, another way of calculating, where the relevance number is converted to a number that corresponds to how good the relevance is compared to what can possibly be achieved. This number can then be compared directly between algorithms.

The minimum and maximum achievable relevance can be seen in Figure 1 on the facing page and in Appendix A. Using $\top_x$ for maximum and $\bot_x$ for minimum at a coverage of x yields the following adjusted relevance:

$$\text{relevance}_{\text{adj}} = \frac{\text{relevance} - \bot_x}{\top_x - \bot_x} \tag{5}$$

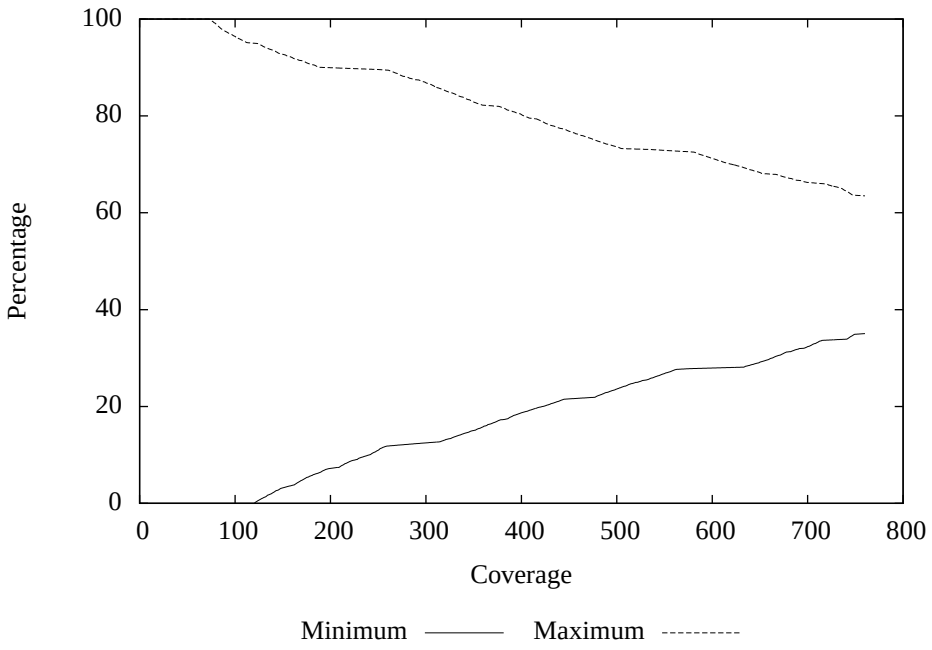

Figure 1: Minimum and maximum values of relevance, given a certain coverage.

2.4 Reporting Relevance Values

We strongly recommend that all relevance values should be given as:

> ... has a relevance of X, or an adjusted relevance of Y at Z coverage,

where X, Y and Z is defined as in Sections 2.1, 2.3 and 2.2, respectively. This ensures that all values will be reported in a similar manner; both brief and immediately comparable to similar reports from other sources.

2.5 Stability Tests

The relative stability of each algorithm reported in Section 4 will be shown using the difference between their original answer set and the answer sets with links removed at various rates, using Spearman Footrule Distance and Spearman ρ rank correlations [Spe04, Spe06, DKNS01], as well as the Relevance Values at the same link removal rates.

3 Files

The dataset consists of a number of files, with very specific content and syntax for each file. This will be further decribed below. All numbers except those given in Section 3.2 are in decimal form.

3.1 Link Files

Three separate link files are supplied, called A.txt, B.txt, and C.txt. The first one contains all hyper-links, the second contains the non-local hyper-links, and the third contains the local hyper-links. All of them share the samt syntax; The first line contains the index of the last web page in the dataset plus one[1] and all other lines contain the numbers corresponding to the start and the end web page of a hyperlink with a horizontal tab inbetween. See Figure 2 for an example.

Please note that while these three files are sorted according to the origin of the hyper-links, they are not further sorted with respect to the recipient.

```
3087532
130     119040
130     1725811
130     1752232
```

Figure 2: The first few lines of the B.txt file.

3.2 Original Page Sets

The files residing in the inv directory contains the hexadecimal index number of each page that contained that stemmed word, i.e. the known dataset (denoted Θ in [Åg08]). The files are DK01, DK03, DK04, EN01, EN02, EN03, EN04, EN05, EN06, EN07, FI01, FI02, FI03, FI04, FI05, FI06, FI07, FI08, FI09, FI10, IS02, IS03, IS06, IS08, IS09, NN01, NN02, NN03, NN04, NN05, NN06, NO01, NO02, NO03, NO04, NO05, NO06, NO07, SE01, SE02, SE03, SE04, SE05, SE06, SE07, SE08, SE09, SE10, SE11, SE12, SE13, SE14, SE15, SE16, SE17, SE18, SE19, SE20, SE21, SE22, SE23, SE24, SE25, SE26, SE27, SE28, SE29, SE30, SE31, SE32, SE33, SE34, SE35, SE36, SE37, and, finally, SE38. A sample file can be seen in Figure 3.

```
13d1b2
1417e6
16bcad
24dc02
```

Figure 3: The first few lines of the inv/SE09 file.

[1] This means that this number can be used for allocating dynamic vectors of sufficient size.

3.3 Assessment Data

These files contain miscellaneous data that were given by or used in the assessment.

The grades.txt file is a colon-separated file with search term (as given in Section 3.2 on the facing page), the web page index, the average grade and the specific grade data as given in Section 2 on page 4. A sample of this file is given in Figure 4.

```
EN01:1172328:0.575000:1:1:10:4:1
EN01:1172329:0.593750:1:1:9:5:1
EN01:1172330:0.593750:1:1:9:5:1
EN01:1194900:0.425000:5:4:5:2:1
```

Figure 4: Sample lines from grades.txt

The spread.txt contains the lower and upper boundary for a certain number of matching pages, as described in Section 2 on page 4. Some sample lines of this file can be seen in Figure 5.

```
1 0.000000 100.000000
2 0.000000 100.000000
          ...
760 35.051752 63.489703
```

Figure 5: Sample lines from spread.txt

4 Sample Results

This section will present the results from some standard algorithms, including straight link count, PageRank, Topic-Sensitive PageRank, HITS, Propagation of Topic-relevance (ProT), and Superpositioned Singleton Propagation of Topic-relevance (S^{2}ProT). The collected relevance results can be seen in Table 1 and Figure 14 on page 15 while the rank order stability can be seen in figures 12 (Spearman Footrule Distance) and 13 on page 14 (Spearman ρ).

4.1 Straight Link Count

Straight Link Count is the total number of incoming links to each page, then selected by 10 highest values per word to search for. Straight Link Count has a relevance of 48.621%, or an adjusted relevance of 46.979% at 73.158% coverage.

4.2 PageRank

Brin and Page [BP98] describe a way to generate a query independent rating for each web page according to the structure of the web. This method is called PageRank. It uses a random surfing model over the Internet, which means that it models a web surfer who randomly follows one link out from the current page or (at a certain probability $1-\mu$) jumps to a random page on the Internet. The original PageRank algorithm gives a value for each page $j \in V$, which is obtained by solving Eq. (6) using iteration to find a fixed point.

$$PR(j) = \frac{1-\mu}{n} + (\mu) \times \sum_{(i,j)\in E} PR(i)/\text{outdegree}(i) \tag{6}$$

Here, n is the number of pages of the web considered and $0 \leq (1-\mu) < 1$ is a damping factor, so that the rating of individual page nests with no outgoing links will not continue to increase in each iteration. This damping factor corresponds to the probability that a user will jump to a random page instead of continuing from the current page.

This is the same as using a matrix P obtained from the column-normalised adjacency matrix M of the web by adding the damping factor:

$$P = \left[\frac{1-\mu}{n}\right]_{n\times n} + \mu M \tag{7}$$

The rating returned, which is called PageRank, is the dominant eigenvector of P: $P\pi = \pi, \pi \geqslant 0, ||\pi||_1 = 1$. This means that the i-th entry of π is the probability that a surfer visits page i, or the PageRank of page i. It has been shown that the second largest eigenvalue of P ($\lambda_2(P)$) will never be larger than μ [HK03], leading to fast convergence when using power iteration to find the PageRanks.[2] While this is sufficient for most applications,there have been a number of proposals for speeding up the calculations [Hav99, HKK$^+$03, KHMG03a, KHMG03b, ANTT01, KHG03, CGR04, LM02, IK06, BLMP06, PCD$^+$08].

[2]Both because the power method converges at a rate proportional to $|\lambda_1/\lambda_2|$ [GV96] and because P is an irreducible n-state Markov chain indicates that power iteration will always converge on a stable value [IM94, Theorem 5.2].

PageRank with $\mu = 0.85$ achieved a relevance of 48.131%, or an adjusted relevance of 46.950% at 89.474% coverage. The relevance stability of PageRank can be seen in Figure 6.

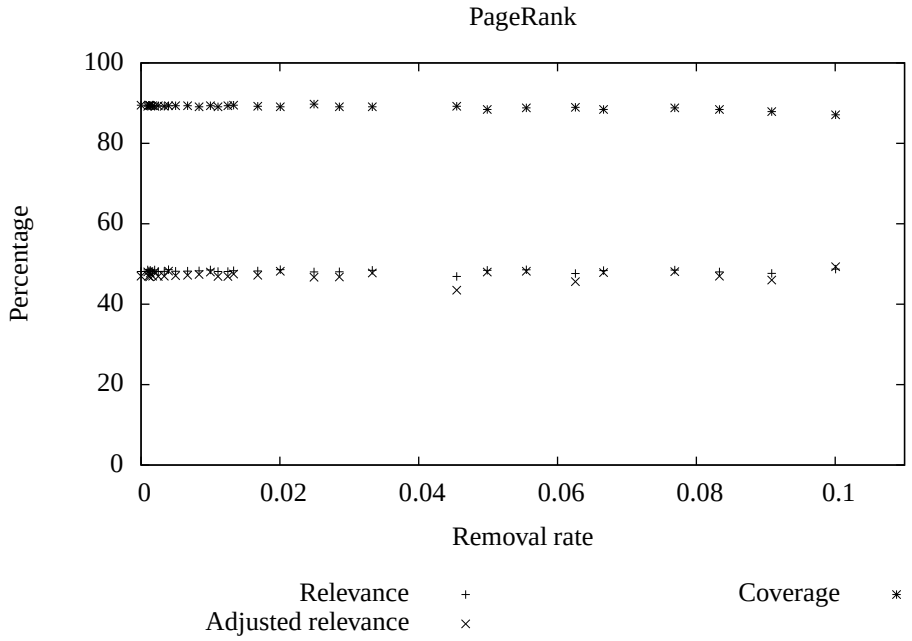

Figure 6: PageRank Relevance when some links have been removed.

4.3 Topic-Sensitive PageRank

A very interesting version of PageRank is Topic-Sensitive PageRank [Hav02]. It uses the same general ideas as the original PageRank, except that the damping factor is not added uniformly. It is scaled and added to a set of pages assumed to be about that particular subject, meaning that the resulting vectors will depend on both the structure of the graph as well as the starting set of pages.

The downside of using this approach is that the resulting matrices are not necessarily irreducible, potentially leading to very slow (or no) convergence when using the power method.

Topic-Sensitive PageRank with $\mu = 0.85$ achieved a relevance of 52.863%, or an adjusted relevance of 59.679% at 93.553% coverage. The relevance stability of Topic-Sensitive PageRank can be seen in Figure 7 on the following page.

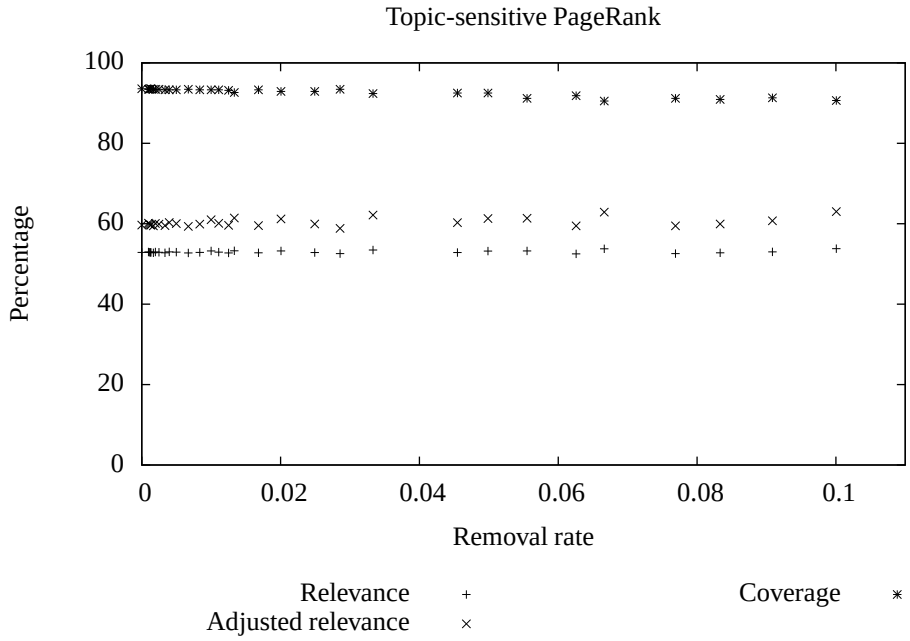

Figure 7: Topic-Sensitive PageRank Relevance when some links have been removed.

4.4 HITS

Kleinberg [Kle98] uses a standard adjacency matrix A, but does not work on the entire Internet directly. HITS requires a bootstrap data set, consisting of pages that are initially assumed to be about a specific subject. This set is further extended with all pages pointed to by the bootstrap set as well as pages that point to the bootstrap set. Each page in the entire set is given a start value in two categories, *hub* (denoting an important link page) and *authority* (denoting a page with valuable information on the given subject). These values are adjusted by simultaneous iteration over the equations given in Eq. (8), where h_i denotes the hub value for page i and a_j the authority value for page j.

$$h_i = \sum_{(i,j)\in E} a_j \qquad a_j = \sum_{(i,j)\in E} h_i \tag{8}$$

HITS achieved a hub relevance of 33.636%, or an adjusted relevance of 33.636% at 0.395% coverage. The corresponding HITS authority relevance was 40.000, or an adjusted relevance of 40.000% at 0.921% coverage. The relevance stability of HITS Authorithy can be seen in Figure 8 on the next page and for HITS Hub in Figure 9 on the facing page.

There are a number of versions of HITS, including CLEVER [CDI98], BHITS [BH98], SALSA [LM01], Randomised HITS [NZJ01], and Subspace HITS [NZJ01]. Results from these will not be given in this paper.

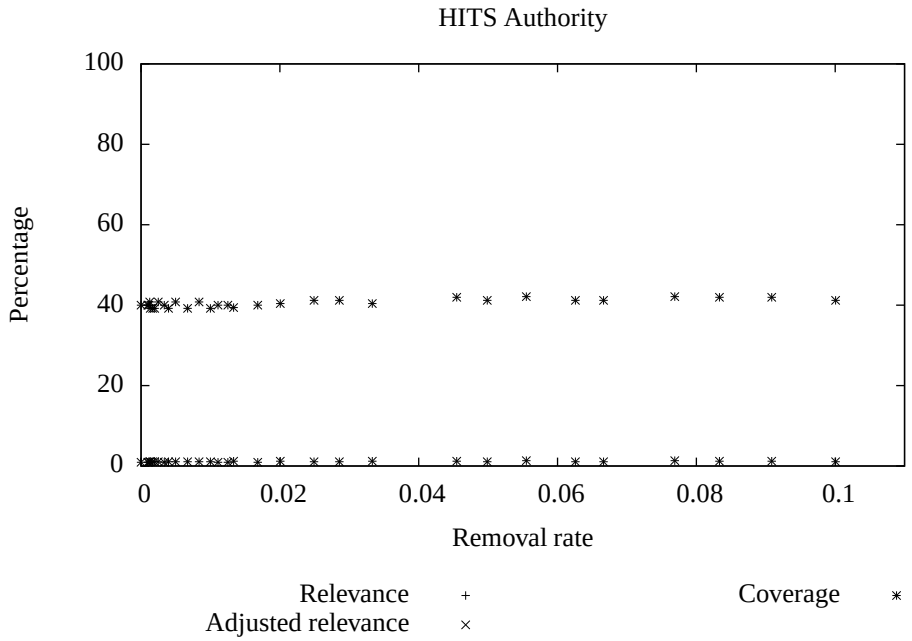

Figure 8: HITS Authority Relevance when some links have been removed.

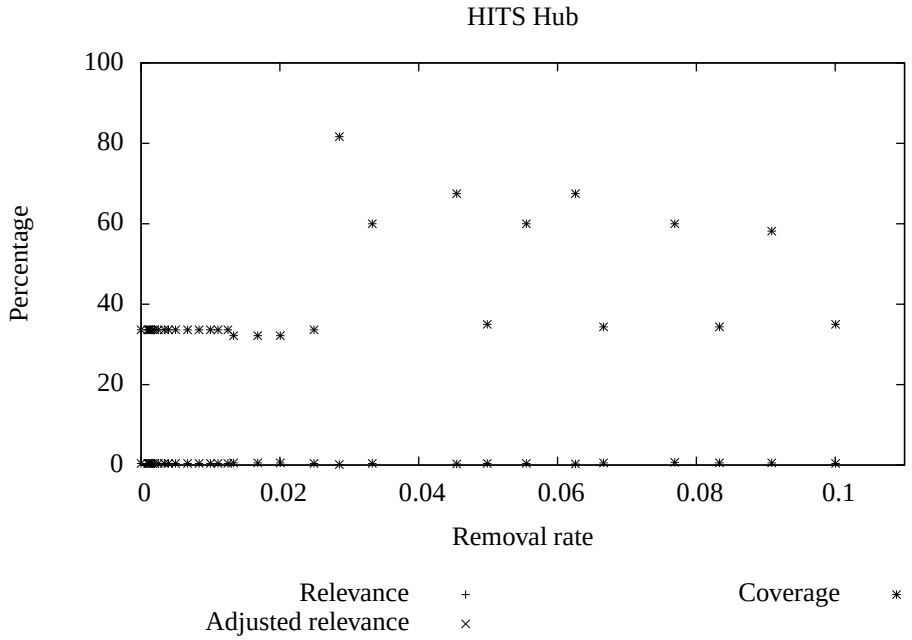

Figure 9: HITS Hub Relevance when some links have been removed.

4.5 ProT: Propagation of Topic-relevance

The basic idea of this algorithm is that each page is assigned a relevance value for each topic. This relevance value propagates along hyperlinks, while decreasing for each link travelled. This decrease is controlled using a parameter that corresponds to a *decay factor*, called ξ [Åg06, Åg08]. The method is quite similar to *spreading activation* [PPR96, Cre97, CL99, AADD05], but uses multiplicative rather than additive activation in each iteration.

It is most easily implemented using an iterative approach. Starting with an initial assignment of relevance values, iterated updates and normalisations are made until a fixed point is reached (or, more precisely, until the changes are smaller than a certain threshold that we call *cut-off*). Let ϖ_j^k be the relevance value for page $j \in V$ in iteration $k \geqslant 0$. The initial values depend on the set of pages initially assumed to be on-topic, i.e., the set Θ:

$$\varpi_j^0 = \begin{cases} 1 & \text{if } j \in \Theta \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

The algorithm is then given by normalisation of

$$\varpi_j^k = \left(\frac{1}{\xi} \sum_{(i,j)\in\text{graph}} \varpi_i^{k-1} \right) + \begin{cases} \varpi_j^{k-1} & \text{if } j \in \Theta \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

This is the same thing as using the wide-spread power method [GV96] on a matrix $\hat{A}$, consisting of a standard adjacency matrix A divided by ξ with the addition that diagonal elements corresponding to pages in the set Θ are set to 1 (and using the values in the diagonal as the starting vector). Each iteration computes the next answer vector ϖ^1, ϖ^2, etc. The final result is given after a suitably large number of iterations (i.e., $\lim_{k\to\infty} \varpi^k = \pi_1(\hat{A})$), but the algorithm usually converges quite quickly as long as normalisation is done after each iteration.

ProT achieved a relevance of 53.995%, or an adjusted relevance of 57.501% at 54.737% coverage. The relevance stability of ProT can be seen in Figure 10 on the next page.

4.6 S^{2}ProT: Superpositioned Singleton Propagation of Topic-relevance

A further development of the ProT algorithm using the same general idea but a slightly different approach is the S^{2}ProT algorithm. Instead of trying to generate the entire eigenvector at once, it creates one vector for each page in Θ and then performs additive superpositioning of these followed by normalisation, thus resulting in the vector that yields the returned rating. The rationale for this is that even though many different calculations need to be performed, this is offset by *much* faster convergence for each subproblem and reuse of vectors whenever a page is on-topic for more than one topic [Åg08].

The reason for the fast propagation is that each such vector calculation can be viewed as a propagation with decreasing strength, i.e. a topological ordering with minor changes because of back links. The convergence of S^{2}ProT was shown in [Åg08].

S^{2}ProT achieved a relevance of 56.528%, or an adjusted relevance of 71.314% at 91.974% coverage. The relevance stability of S^{2}ProT can be seen in Figure 11 on the facing page.

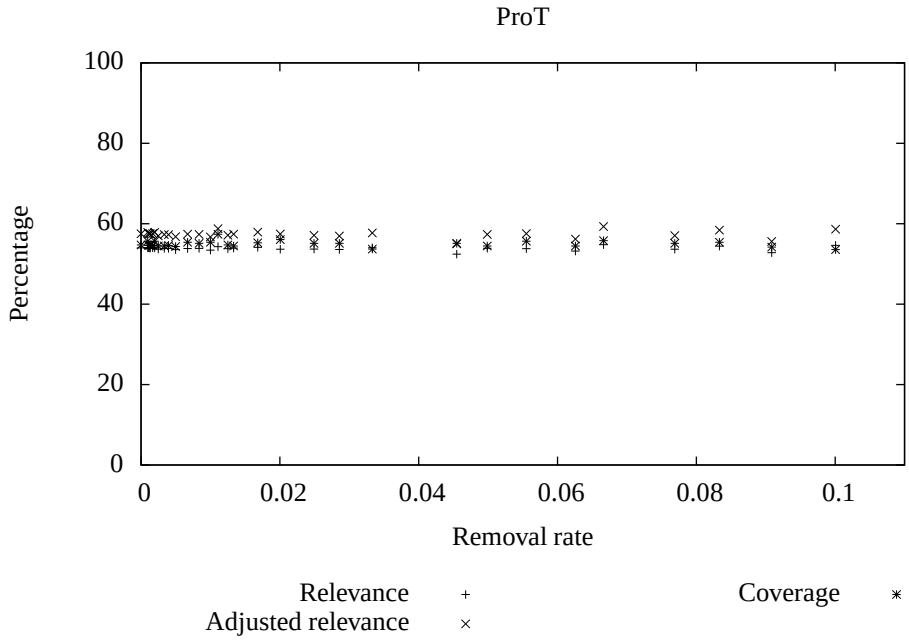

Figure 10: ProT Relevance when some links have been removed.

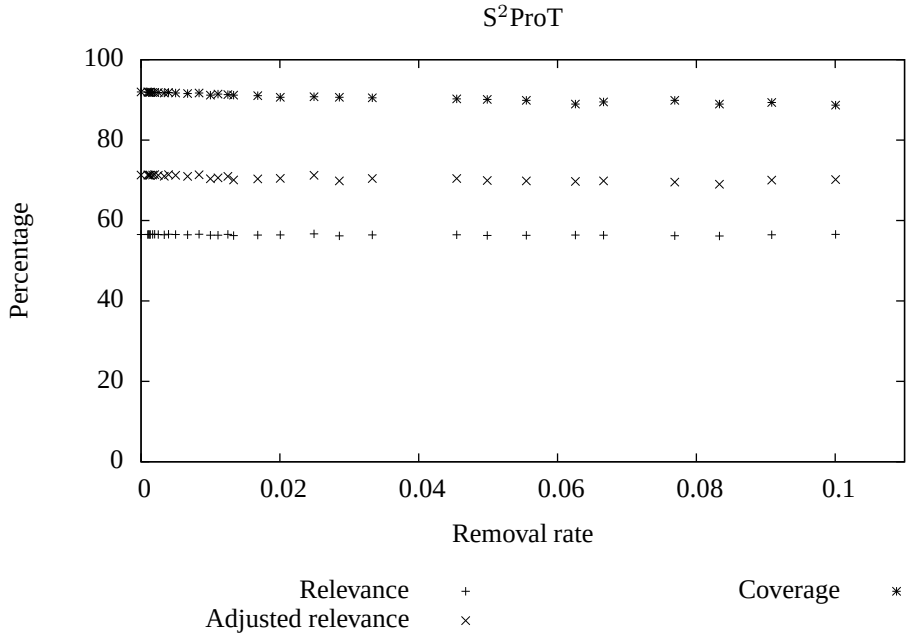

Figure 11: S^2ProT Relevance when some links have been removed.

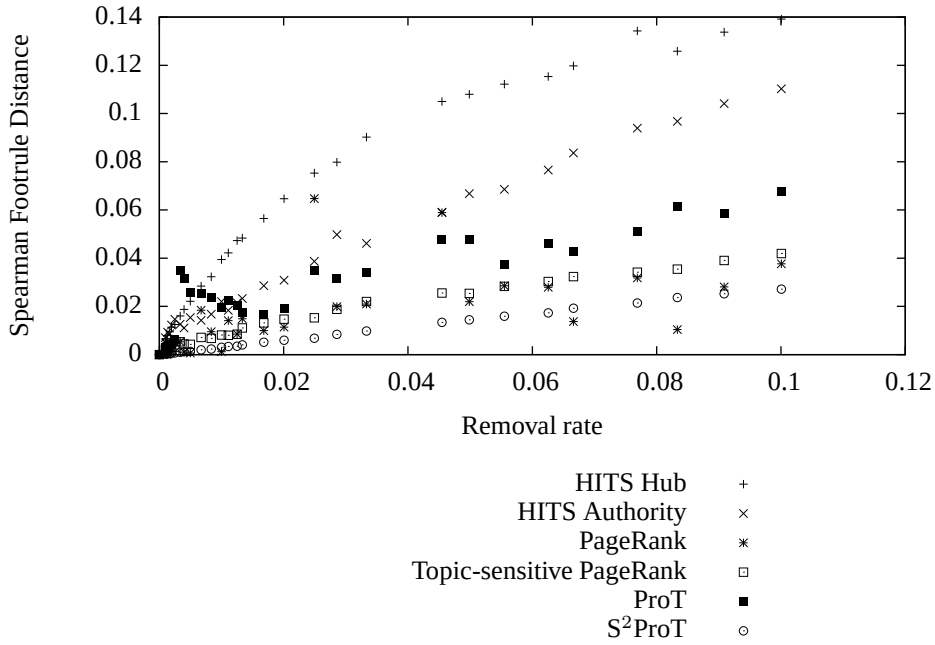

Figure 12: Spearman Footrule Distance for ranking order when some links have been removed.

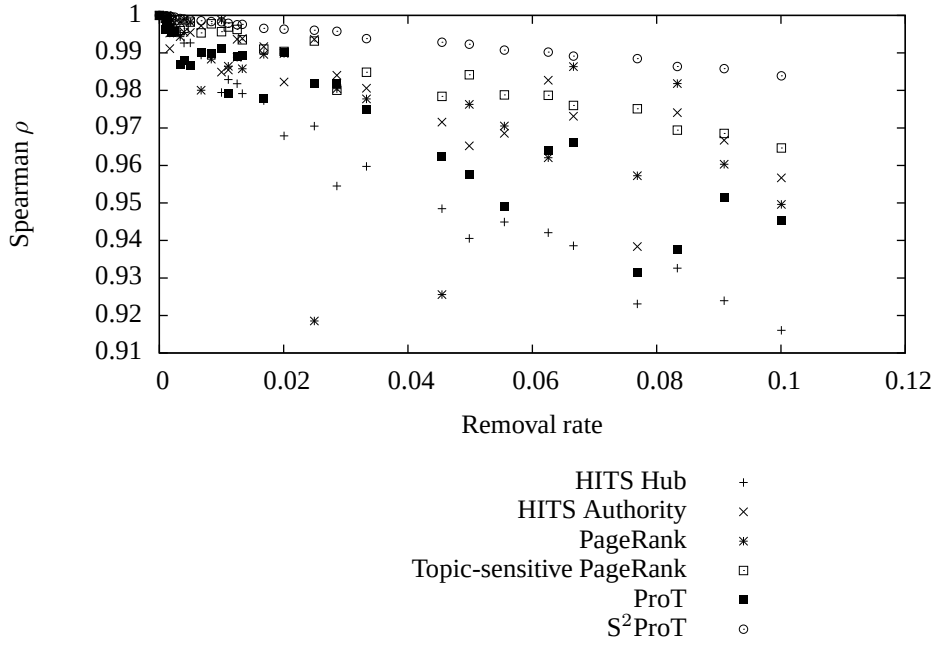

Figure 13: Spearman ρ for ranking order when some links have been removed.

Table 1: Relevance values for some common web-search algorithms

Algorithm	Relevance	Adj. Rel.	Coverage	σ	Count
Straight Link Count	48.621%	46.976%	73.158%	37.108%	3525
HITS Hub	33.636%	33.636%	0.395%	37.359%	22
HITS Authority	40.000%	40.000%	0.921%	29.368%	49
PageRank	48.131%	46.950%	89.474%	36.931%	4419
Topic-Sensitive PageRank	52.863%	59.679%	93.553%	36.647%	4747
ProT	53.995%	57.501%	54.737%	36.322%	2861
S^2ProT	56.528%	71.314%	91.974%	35.937%	4631

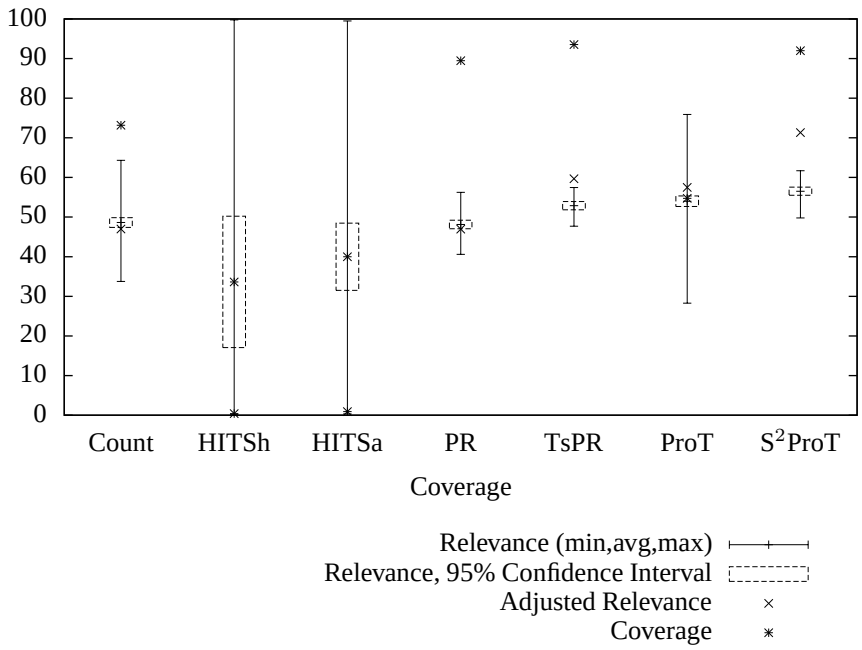

Figure 14: The collected results, where Count is Straight Link Count, HITSh is HITS Hub, HITSa is HITS Authority, PR is PageRank, TsPR is Topic-Sensitive PageRank, ProT is Propagation of Topic-relevance, and S^2ProT is Superpositioned Singleton Propagation of Topic-relevance. The min and max values constitutes the extreme values possible including the misses set to all 0 or 1 (as in equations 2 and 3 on page 4). Please note that we have used Student's t-distribution for the confidence intervals for HITS and that the Adjusted Relevance can be outside of the Confidence Interval for the Relevance.

5 Discussion

The dataset is derived from real world data and large enough to be used as basis for research and assessment of web search engines. Moreover, the dataset has been assessed in a fashion that was as objective as possible (with details given in [Åg08]), so as to not let any personal bias shift the data towards the results from any specific search algorithm. This means that the assessment data given is very reliable and trustworthy.

Note that the dataset appears to contain much less spam links[3] than what's expected among the commercial web sites using *.com* addresses, except when looking at the local links on one of the major Swedish newspaper sites. This means that it might be slightly easier to obtain good search results on this dataset than on others.

When looking at stability of the algorithms, it is obvious that HITS, PageRank and ProT are somewhat sensitive to perturbed data, small changes (in this case link removals) can greatly affect the results. Topic-Sensitive PageRank and S^2ProT are on the other hand very stable. Hindsight indicates that it might have been a good idea to add one or more of the updated versions of HITS (e.g. Randomised HITS [NZJ01]) to the collection of algorithms to test, but that had to be left to a later date.

We have left one major task for future works; to assess the standard information retrieval statistics (such as number of each word appearing in each file and inverse document frequency) of the dataset.

The dataset is available for download from `http://www8.cs.umu.se/~ola/ProT/`.

References

[AADD05] Dipti Aswath, Syed Toufeeq Ahmed, James D'cunha, and Hasan Davulcu. Boosting Item Keyword Search with Spreading Activation. In Andrzej Skowron, Rakesh Agrawal, Michael Luck, Takahira Yamaguchi, Pierre Morizet-Mahoudeaux, Jiming Liu, and Ning Zhong, editors, *Web Intelligence*, pages 704–707. IEEE Computer Society, 2005.

[ANTT01] A. Arasu, J. Novak, A. Tomkins, and J. Tomlin. PageRank computation and the structure of the web: Experiments and algorithms. Technical report, IBM Almaden Research Center, November 2001.

[BH98] Krishna Bharat and Monika R. Henzinger. Improved algorithms for topic distillation in a hyperlinked environment. In *SIGIR '98: Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 104–111, New York, NY, USA, 1998. ACM Press.

[BLMP06] A. Z. Broder, R. Lempel, F. Maghoul, and J. Pedersen. Efficient PageRank approximation via graph aggregation. *Information Retrieval*, 9(2):123–138, March 2006.

[BP98] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117, 1998.

[3] Links to increase PageRank values and thus rank orders of web sites.

[CDI98] Soumen Chakrabarti, Byron E. Dom, and Piotr Indyk. Enhanced hypertext categorization using hyperlinks. In Laura M. Haas and Ashutosh Tiwary, editors, *Proceedings of SIGMOD-98, ACM International Conference on Management of Data*, pages 307–318, Seattle, US, 1998. ACM Press, New York, US.

[CGR04] Gianna M. Del Corso, Antonio Gulli, and Francesco Romani. Fast PageRank computation via a sparse linear system. In *Proceedings of Third Workshop on Algorithms and Models for the Web-Graph (WAW 2004)*, Rome, Italy, October 16, 2004.

[CL99] Fabio Crestani and Puay Leng Lee. WebSCSA: Web Search by Constrained Spreading Activation. In *IEEE Forum on Research and Technology Advances in Digital Libraries (ADL '99)*, pages 163–170, 1999.

[Cre97] Fabio Crestani. Application of Spreading Activation Techniques in Information Retrieval. *Artif. Intell. Rev.*, 11(6):453–482, 1997.

[DKNS01] Cynthia Dwork, Ravi Kumar, Moni Naor, and D. Sivakumar. Rank aggregation methods for the web. In *WWW '01: Proceedings of the 10th international conference on World Wide Web*, pages 613–622, New York, NY, USA, 2001. ACM Press.

[GV96] Gene H. Golub and Charles F. Van Loan. *Matrix computations (3rd ed.)*. Johns Hopkins University Press, 1996.

[Hav99] Taher H. Haveliwala. Efficient computation of PageRank. Technical Report 1999-31, Stanford University Database Group, October 18, 1999.

[Hav02] Taher H. Haveliwala. Topic-sensitive PageRank. In *Proceedings of the eleventh international conference on World Wide Web*, pages 517–526. ACM Press, 2002.

[HK03] Taher H. Haveliwala and Sepandar D. Kamvar. The second eigenvalue of the google matrix. Technical report, Stanford University, March 2003.

[HKK+03] Taher Haveliwala, Sepandar Kamvar, Dan Klein, Chris Manning, and Gene Golub. Computing PageRank using power extrapolation. Technical report, Stanford University, CA, USA, October 18, 2003.

[IK06] Ilse C. F. Ipsen and Steven Kirkland. Convergence analysis of a PageRank updating algorithm by Langville and Meyer. *SIAM J. Matrix Anal. Appl.*, 27(4):952–967, 2006.

[IM94] Ilse C. F. Ipsen and Carl D. Meyer. Uniform stability of markov chains. *SIAM J. Matrix Anal. Appl.*, 15(4):1061–1074, 1994.

[KHG03] Sepandar D. Kamvar, Taher H. Haveliwala, and Gene H. Golub. Adaptive methods for the computation of PageRank. Technical report, Stanford University, CA, USA, April 2003.

[KHMG03a] Sepandar D. Kamvar, Taher H. Haveliwala, Christopher D. Manning, and Gene H. Golub. Exploiting the block structure of the web for computing PageRank. Technical report, Stanford University, CA, USA, March 4, 2003.

[KHMG03b] Sepandar D. Kamvar, Taher H. Haveliwala, Christopher D. Manning, and Gene H. Golub. Extrapolation methods for accelerating PageRank computations. In *Proceedings of the Twelfth International World Wide Web Conference*, 2003.

[Kle98] J. Kleinberg. Authoritative sources in a hyperlinked environment. In *Proc. of ACM-SIAM Symposium on Discrete Algorithms*, pages 668–677, 1998.

[LM01] R. Lempel and S. Moran. SALSA: The stochastic approach for link-structure analysis. *ACM Trans. Inf. Syst.*, 19(2):131–160, 2001.

[LM02] Amy N. Langville and Carl D. Meyer. Updating PageRank using the group inverse and stochastic complementation. Technical Report CRSC-TR02-32, Center for Research in Scientific Computation, North Carolina State University, Raleigh, NC, USA, November 2002.

[NZJ01] Andrew Y. Ng, Alice X. Zheng, and Michael Jordan. Stable algorithms for link analysis. In *Proceedings of the Twenty-fourth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, September 2001.

[PCD+08] Josiane Xavier Parreira, Carlos Castillo, Debora Donato, Sebastian Michel, and Gerhard Weikum. The Juxtaposed approximate PageRank method for robust PageRank approximation in a peer-to-peer web search network. *The International Journal on Very Large Data Bases*, 17(2):291–313, March 2008.

[PPR96] Peter Pirolli, James E. Pitkow, and Ramana Rao. Silk from a Sow's Ear: Extracting Usable Structures from the Web. In *CHI*, pages 118–125, 1996.

[Spe04] Charles Spearman. The proof and measurement of association between two things. *American Journal of Psychology*, 15:72–101, January 1904.

[Spe06] Charles Spearman. 'Footrule' for measuring correlations. *British Journal of Psychology*, 2:89–108, June 1906.

[Åg06] Ola Ågren. Assessment of WWW-Based Ranking Systems for Smaller Web Sites. *INFOCOMP Journal of Computer Science*, 5(2):45–55, June 2006.

[Åg08] Ola Ågren. S^2ProT: Rank Allocation by Superpositioned Propagation of Topic-Relevance. *International Journal of Web Information Systems*, 4(4):416–440, 2008.

A Maximum Spread of Relevance

Table 2: Minimum and maximum possible relevance numbers

x	**Minimum**	**Maximum**	x	**Minimum**	**Maximum**
1	0.000000	100.000000	103	0.000000	96.050420
...	...	...	104	0.000000	95.954825
73	0.000000	100.000000	105	0.000000	95.902637
74	0.000000	99.836066	106	0.000000	95.767717
75	0.000000	99.696970	107	0.000000	95.675676
76	0.000000	99.571429	108	0.000000	95.576560
77	0.000000	99.459459	109	0.000000	95.481481
78	0.000000	99.268293	110	0.000000	95.390200
79	0.000000	99.171598	111	0.000000	95.283688
80	0.000000	98.994413	112	0.000000	95.146299
81	0.000000	98.835979	113	0.000000	95.128645
82	0.000000	98.620690	114	0.000000	95.111111
83	0.000000	98.433180	115	0.000000	95.093697
84	0.000000	98.268398	116	0.000000	95.076401
85	0.000000	98.064516	117	0.000000	95.059222
86	0.000000	97.923077	118	0.000000	95.042159
87	0.000000	97.794118	119	0.000000	95.025210
88	0.000000	97.676056	120	0.000000	95.008375
89	0.000000	97.567568	121	0.152439	94.991653
90	0.000000	97.443366	122	0.293255	94.975042
91	0.000000	97.343750	123	0.424929	94.958541
92	0.000000	97.250755	124	0.547945	94.942149
93	0.000000	97.118156	125	0.664894	94.846029
94	0.000000	97.008310	126	0.775194	94.746032
95	0.000000	96.927224	127	0.881612	94.642857
96	0.000000	96.850394	128	0.982801	94.544073
97	0.000000	96.760204	129	1.081731	94.410029
98	0.000000	96.601942	130	1.176471	94.306358
99	0.000000	96.494118	131	1.267281	94.212766
100	0.000000	96.413793	132	1.357466	94.122563
101	0.000000	96.261062	133	1.497797	94.035568
102	0.000000	96.137339	134	1.666667	93.956931

Continued on next page

Table 2 – continued from previous page

x	**Minimum**	**Maximum**	x	**Minimum**	**Maximum**
135	1.747368	93.880795	172	4.974874	91.257750
136	1.825726	93.807040	173	5.086634	91.185250
137	1.901840	93.740360	174	5.195122	91.064014
138	1.975806	93.675539	175	5.341317	90.965812
139	2.121807	93.612500	176	5.398335	90.889077
140	2.235067	93.538841	177	5.454545	90.813758
141	2.344045	93.446602	178	5.581395	90.758333
142	2.476895	93.389423	179	5.651672	90.703642
143	2.603978	93.251174	180	5.720824	90.667766
144	2.665474	93.130035	181	5.788876	90.588235
145	2.725664	93.013544	182	5.921788	90.544715
146	2.847222	92.925473	183	5.986696	90.493129
147	2.964225	92.839912	184	6.050605	90.417001
148	3.082077	92.819672	185	6.113537	90.317460
149	3.139535	92.799564	186	6.175515	90.195925
150	3.196046	92.740022	187	6.236559	90.054054
151	3.251634	92.684492	188	6.296692	90.046296
152	3.306321	92.629905	189	6.388596	90.038551
153	3.360129	92.563025	190	6.478579	90.030817
154	3.413078	92.432990	191	6.597938	90.023095
155	3.465190	92.339776	192	6.714140	90.015385
156	3.516484	92.295248	193	6.791120	90.007686
157	3.566978	92.251256	194	6.896208	90.000000
158	3.616692	92.193420	195	6.983185	89.992325
159	3.665644	92.110454	196	7.040315	89.984663
160	3.713851	92.043011	197	7.096774	89.977011
161	3.761329	91.900192	198	7.152575	89.969372
162	3.808096	91.825095	199	7.180233	89.961744
163	3.950073	91.764151	200	7.207729	89.954128
164	4.086331	91.704120	201	7.235067	89.946524
165	4.217207	91.644981	202	7.262248	89.938931
166	4.327323	91.586716	203	7.289272	89.931350
167	4.433834	91.484517	204	7.316141	89.923780
168	4.536913	91.462307	205	7.342857	89.916222
169	4.692005	91.440217	206	7.369421	89.908676
170	4.753247	91.418248	207	7.395833	89.901141
171	4.865900	91.331546	208	7.422096	89.893617

Continued on next page

Table 2 – continued from previous page

x	**Minimum**	**Maximum**	x	**Minimum**	**Maximum**
209	7.448211	89.886105	246	10.542857	89.615953
210	7.585887	89.878604	247	10.631206	89.608856
211	7.719780	89.871114	248	10.764912	89.601770
212	7.813067	89.863636	249	10.869868	89.594694
213	7.915544	89.856170	250	10.966851	89.587629
214	8.016014	89.848714	251	11.106557	89.573529
215	8.166667	89.841270	252	11.287062	89.559471
216	8.240418	89.833837	253	11.378849	89.545455
217	8.313149	89.826415	254	11.469415	89.531479
218	8.424658	89.819005	255	11.558785	89.517544
219	8.522920	89.811605	256	11.629435	89.503650
220	8.619529	89.804217	257	11.699346	89.489796
221	8.714524	89.796840	258	11.751302	89.475983
222	8.798674	89.789474	259	11.802853	89.462209
223	8.844884	89.782119	260	11.836999	89.448476
224	8.890715	89.774775	261	11.854005	89.427951
225	8.936170	89.767442	262	11.870968	89.353448
226	8.981255	89.760120	263	11.887887	89.280114
227	9.025974	89.752809	264	11.904762	89.194915
228	9.070331	89.745509	265	11.921594	89.110644
229	9.217877	89.738220	266	11.938383	89.027778
230	9.328063	89.730942	267	11.955128	88.952447
231	9.394654	89.723674	268	11.971831	88.878249
232	9.460516	89.716418	269	11.988491	88.823129
233	9.525661	89.709172	270	12.005109	88.750000
234	9.590101	89.701937	271	12.021684	88.647925
235	9.653846	89.694713	272	12.038217	88.518519
236	9.716909	89.687500	273	12.054707	88.426230
237	9.779300	89.680297	274	12.071156	88.340924
238	9.841030	89.673105	275	12.087563	88.262274
239	9.902108	89.665924	276	12.103929	88.189987
240	9.962547	89.658754	277	12.120253	88.167841
241	10.022355	89.651594	278	12.136536	88.145780
242	10.081542	89.644444	279	12.152778	88.101911
243	10.255848	89.637306	280	12.168979	88.058376
244	10.355330	89.630178	281	12.185139	87.972292
245	10.453237	89.623060	282	12.201258	87.892433

Continued on next page

Table 2 – continued from previous page

x	**Minimum**	**Maximum**	x	**Minimum**	**Maximum**
283	12.217337	87.777090	320	13.079179	85.244399
284	12.233375	87.740741	321	13.182083	85.164557
285	12.249373	87.688998	322	13.232908	85.085599
286	12.265332	87.637699	323	13.283324	85.007511
287	12.281250	87.571516	324	13.333333	84.893194
288	12.297129	87.490931	325	13.369628	84.871032
289	12.312968	87.475845	326	13.392449	84.834240
290	12.328767	87.460796	327	13.501703	84.783037
291	12.344527	87.445783	328	13.587140	84.717722
292	12.360248	87.430806	329	13.658810	84.638672
293	12.375931	87.386091	330	13.729760	84.560466
294	12.391574	87.312277	331	13.800000	84.469112
295	12.407178	87.224852	332	13.869541	84.337637
296	12.422744	87.142857	333	13.938394	84.230769
297	12.438272	87.090164	334	14.006568	84.145420
298	12.453761	87.037901	335	14.074074	84.074248
299	12.469212	86.986063	336	14.140921	84.003742
300	12.484625	86.846640	337	14.207120	83.953380
301	12.500000	86.759840	338	14.281116	83.909809
302	12.515337	86.723549	339	14.354322	83.866481
303	12.530637	86.664779	340	14.443266	83.780544
304	12.545900	86.548822	341	14.495510	83.708123
305	12.561125	86.504474	342	14.547368	83.560709
306	12.576313	86.416667	343	14.598846	83.450768
307	12.591463	86.299559	344	14.649948	83.399370
308	12.606577	86.226725	345	14.716589	83.348294
309	12.621655	86.104178	346	14.806202	83.297539
310	12.636695	86.004308	347	14.899846	83.174956
311	12.651699	85.926124	348	14.953941	83.054209
312	12.666667	85.820499	349	15.007649	82.962314
313	12.681598	85.792263	350	15.045778	82.871460
314	12.696493	85.719557	351	15.091278	82.781629
315	12.711353	85.647614	352	15.128983	82.739844
316	12.795441	85.576421	353	15.253387	82.656922
317	12.878427	85.497925	354	15.332002	82.574850
318	12.946058	85.420320	355	15.409836	82.498936
319	13.012972	85.335725	356	15.479723	82.463891

Continued on next page

Table 2 – continued from previous page

x	**Minimum**	**Maximum**	x	**Minimum**	**Maximum**
357	15.541872	82.384778	394	18.268110	80.702495
358	15.643661	82.306397	395	18.328494	80.619029
359	15.730282	82.228739	396	18.402147	80.555979
360	15.815981	82.214316	397	18.475134	80.520715
361	15.939452	82.199916	398	18.547463	80.485584
362	15.971223	82.185541	399	18.619145	80.408009
363	16.034400	82.171190	400	18.690187	80.281426
364	16.121673	82.156863	401	18.756058	80.156541
365	16.262959	82.142559	402	18.821400	80.066766
366	16.317520	82.128280	403	18.873296	79.977852
367	16.371723	82.114024	404	18.924860	79.908054
368	16.425572	82.099792	405	18.963317	79.838769
369	16.502555	82.085584	406	19.001591	79.762774
370	16.578826	82.071399	407	19.085873	79.680233
371	16.654395	82.057238	408	19.169291	79.563019
372	16.729272	82.043100	409	19.239514	79.541847
373	16.803466	82.028986	410	19.284597	79.520721
374	16.876986	82.014894	411	19.361868	79.499640
375	16.994142	82.000827	412	19.426357	79.478605
376	17.103850	81.986782	413	19.490347	79.457615
377	17.168970	81.944788	414	19.553846	79.436670
378	17.233570	81.889117	415	19.616858	79.401862
379	17.254989	81.806293	416	19.675449	79.360257
380	17.276351	81.729048	417	19.733638	79.291311
381	17.297656	81.616815	418	19.791430	79.216034
382	17.318905	81.567916	419	19.848828	79.134581
383	17.361233	81.484156	420	19.894459	79.020365
384	17.403339	81.379585	421	19.939850	78.980378
385	17.445223	81.238133	422	19.985002	78.851375
386	17.486888	81.195738	423	20.029918	78.737024
387	17.625650	81.153543	424	20.074599	78.630420
388	17.756768	81.111548	425	20.144874	78.524983
389	17.846154	81.048924	426	20.214576	78.420694
390	17.934552	80.986739	427	20.283714	78.342373
391	18.036318	80.896391	428	20.334313	78.264686
392	18.145973	80.847523	429	20.416819	78.187626
393	18.207271	80.798919	430	20.473761	78.089662

Continued on next page

Table 2 – continued from previous page

x	**Minimum**	**Maximum**	x	**Minimum**	**Maximum**
431	20.562001	78.044059	468	21.794177	75.658479
432	20.625000	77.998668	469	21.806495	75.577830
433	20.687545	77.953488	470	21.818796	75.509844
434	20.749641	77.908518	471	21.831081	75.454279
435	20.811294	77.830967	472	21.843349	75.410939
436	20.879316	77.754028	473	21.855601	75.336637
437	20.953563	77.677694	474	21.867835	75.274629
438	21.003888	77.601958	475	21.880054	75.151690
439	21.073944	77.532510	476	21.892256	75.083525
440	21.143458	77.469216	477	21.904441	75.015791
441	21.225559	77.411955	478	22.016722	74.948483
442	21.306917	77.386254	479	22.109817	74.881598
443	21.384562	77.360619	480	22.201987	74.808075
444	21.461565	77.309547	481	22.287409	74.735052
445	21.506190	77.258737	482	22.348883	74.680611
446	21.518900	77.183009	483	22.409954	74.608451
447	21.531593	77.107859	484	22.470627	74.529891
448	21.544269	77.008547	485	22.590791	74.469633
449	21.556927	76.910236	486	22.639069	74.409713
450	21.569568	76.842436	487	22.734878	74.350125
451	21.582192	76.780350	488	22.829706	74.266704
452	21.594798	76.718653	489	22.876756	74.201216
453	21.607387	76.657338	490	22.923567	74.177250
454	21.619959	76.559406	491	22.970140	74.129477
455	21.632514	76.504165	492	23.051866	74.068681
456	21.645051	76.449231	493	23.127559	74.008219
457	21.657572	76.327217	494	23.197364	73.971554
458	21.670075	76.296183	495	23.261413	73.898582
459	21.682561	76.234399	496	23.325023	73.826087
460	21.695031	76.129618	497	23.388199	73.783564
461	21.707483	76.056126	498	23.445820	73.734777
462	21.719918	76.013249	499	23.544811	73.679806
463	21.732337	75.970553	500	23.614606	73.625101
464	21.744739	75.928036	501	23.683889	73.570661
465	21.757123	75.860729	502	23.752665	73.510182
466	21.769492	75.793864	503	23.843589	73.390192
467	21.781843	75.732106	504	23.883848	73.324475

Continued on next page

Table 2 – continued from previous page

x	**Minimum**	**Maximum**	x	**Minimum**	**Maximum**
505	23.981342	73.252980	542	26.253362	72.985075
506	24.055755	73.246822	543	26.321716	72.973044
507	24.124851	73.240667	544	26.367041	72.961025
508	24.193452	73.234516	545	26.440949	72.949020
509	24.210526	73.228367	546	26.520584	72.937026
510	24.298401	73.222222	547	26.612007	72.925046
511	24.385500	73.216080	548	26.674578	72.913078
512	24.483871	73.209942	549	26.736787	72.901122
513	24.552893	73.203807	550	26.798637	72.889179
514	24.621433	73.197674	551	26.911841	72.877248
515	24.701276	73.191546	552	26.951315	72.865330
516	24.740955	73.185420	553	27.002597	72.853424
517	24.780474	73.179298	554	27.059585	72.841530
518	24.854676	73.173178	555	27.116279	72.829649
519	24.912482	73.167063	556	27.184441	72.817780
520	24.969957	73.160950	557	27.258023	72.805924
521	24.992855	73.154840	558	27.336915	72.794079
522	25.015701	73.148734	559	27.420999	72.782248
523	25.102331	73.142631	560	27.512690	72.770428
524	25.154522	73.136531	561	27.592312	72.758621
525	25.206448	73.130435	562	27.631911	72.746826
526	25.287324	73.124341	563	27.665742	72.735043
527	25.356541	73.118251	564	27.676997	72.723272
528	25.385478	73.112164	565	27.688240	72.711514
529	25.414334	73.106081	566	27.699471	72.699767
530	25.443109	73.100000	567	27.710692	72.688033
531	25.471803	73.093923	568	27.721901	72.676311
532	25.528953	73.087849	569	27.733099	72.664601
533	25.592562	73.081778	570	27.744285	72.658751
534	25.655783	73.075710	571	27.755461	72.652903
535	25.770501	73.069645	572	27.766625	72.647059
536	25.856399	73.057526	573	27.777778	72.635379
537	25.890561	73.045419	574	27.788920	72.623711
538	25.931185	73.033325	575	27.800050	72.612056
539	26.035326	73.021243	576	27.811170	72.600412
540	26.098618	73.009174	577	27.816725	72.588780
541	26.184566	72.997118	578	27.822278	72.571355

Continued on next page

Table 2 – continued from previous page

x	**Minimum**	**Maximum**	x	**Minimum**	**Maximum**
579	27.827828	72.553957	616	28.031242	70.232612
580	27.833375	72.536585	617	28.036688	70.164695
581	27.838919	72.501923	618	28.042131	70.125058
582	27.844461	72.467366	619	28.047572	70.073954
583	27.850000	72.387146	620	28.053010	70.023063
584	27.855536	72.296165	621	28.058445	69.972382
585	27.861069	72.209008	622	28.063877	69.921911
586	27.866600	72.128089	623	28.069307	69.899036
587	27.872128	72.053280	624	28.074734	69.842033
588	27.877653	71.984465	625	28.080158	69.785290
589	27.883175	71.927018	626	28.085580	69.710772
590	27.888695	71.869858	627	28.090999	69.670305
591	27.894212	71.823837	628	28.096415	69.629966
592	27.899726	71.783408	629	28.101829	69.560788
593	27.905237	71.743119	630	28.107240	69.474278
594	27.910745	71.628423	631	28.112648	69.456595
595	27.916251	71.554353	632	28.118054	69.403690
596	27.921754	71.491293	633	28.123457	69.333483
597	27.927255	71.428571	634	28.190547	69.257104
598	27.932752	71.366187	635	28.302303	69.209292
599	27.938247	71.286166	636	28.341893	69.137893
600	27.943739	71.234837	637	28.381348	69.066874
601	27.949228	71.183733	638	28.449367	68.972770
602	27.954715	71.110039	639	28.483228	68.919576
603	27.960199	71.064547	640	28.550654	68.866593
604	27.965680	70.979122	641	28.607350	68.813820
605	27.971159	70.916707	642	28.663772	68.784348
606	27.976634	70.815062	643	28.709677	68.719298
607	27.982107	70.775636	644	28.780605	68.642029
608	27.987578	70.736342	645	28.833533	68.594193
609	27.993045	70.697178	646	28.886233	68.546524
610	27.998510	70.612004	647	28.956149	68.499021
611	28.003972	70.498589	648	28.973565	68.439331
612	28.009432	70.436312	649	29.038005	68.379900
613	28.014888	70.374357	650	29.156028	68.296264
614	28.020342	70.345875	651	29.214623	68.207060
615	28.025794	70.289112	652	29.285042	68.100343

Continued on next page

Table 2 – continued from previous page

x	**Minimum**	**Maximum**	x	**Minimum**	**Maximum**
653	29.326133	68.088298	690	31.845494	66.695972
654	29.379100	68.076264	691	31.948663	66.675521
655	29.443795	68.064240	692	31.962797	66.655095
656	29.515265	68.052226	693	31.976918	66.634694
657	29.581590	68.040223	694	31.991028	66.579055
658	29.624043	68.028229	695	32.005125	66.481105
659	29.696829	68.016246	696	32.019210	66.427267
660	29.799447	68.004274	697	32.097530	66.386299
661	29.921983	67.992311	698	32.175483	66.345414
662	29.977101	67.980359	699	32.263191	66.287909
663	30.031993	67.968417	700	32.358411	66.271221
664	30.100296	67.956485	701	32.409791	66.254545
665	30.199864	67.944563	702	32.461020	66.237884
666	30.298710	67.932651	703	32.570888	66.221235
667	30.367946	67.902917	704	32.660800	66.204600
668	30.443494	67.849554	705	32.808848	66.187979
669	30.518635	67.796394	706	32.878567	66.171371
670	30.562654	67.731697	707	32.978812	66.154776
671	30.606535	67.661461	708	33.001868	66.138195
672	30.683036	67.620653	709	33.090307	66.121627
673	30.785142	67.543085	710	33.205287	66.105072
674	30.886525	67.465940	711	33.287304	66.088531
675	30.987191	67.417870	712	33.446931	66.072003
676	31.093509	67.367871	713	33.507585	66.055489
677	31.169629	67.298875	714	33.534112	66.038987
678	31.245349	67.254943	715	33.646025	66.022499
679	31.264217	67.211139	716	33.664964	66.006024
680	31.283060	67.167462	717	33.674428	65.989562
681	31.301879	67.123912	718	33.683888	65.956679
682	31.320672	67.091699	719	33.693344	65.923848
683	31.397779	67.059553	720	33.702796	65.891069
684	31.470460	66.974634	721	33.712245	65.842000
685	31.530457	66.921968	722	33.721689	65.723973
686	31.610811	66.869476	723	33.731130	65.671315
687	31.705109	66.796555	724	33.740567	65.618782
688	31.752045	66.736950	725	33.750000	65.557322
689	31.798839	66.716448	726	33.759429	65.521144

Continued on next page

Table 2 – concluded from previous page

x	**Minimum**	**Maximum**	x	**Minimum**	**Maximum**
727	33.768854	65.476002	744	34.294936	64.063584
728	33.778276	65.430949	745	34.414251	63.932347
729	33.787694	65.385986	746	34.533039	63.789534
730	33.797107	65.341111	747	34.653307	63.647486
731	33.806517	65.296326	748	34.773045	63.635321
732	33.815923	65.231801	749	34.896249	63.623161
733	33.825326	65.101318	750	34.922218	63.611005
734	33.834724	65.072749	751	34.935194	63.598854
735	33.844119	65.044213	752	34.948166	63.586707
736	33.853510	64.987242	753	34.961132	63.574566
737	33.862897	64.854179	754	34.974093	63.562428
738	33.872280	64.728622	755	34.987049	63.550296
739	33.881659	64.603700	756	35.000000	63.538168
740	33.891035	64.535200	757	35.012946	63.526045
741	33.900407	64.466887	758	35.025886	63.513926
742	34.017441	64.330815	759	35.038821	63.501812
743	34.156553	64.195480	760	35.051752	63.489703

Index

Θ, 6, 12
λ, 3, 8
μ, 8
π, 8
ξ, 12

adjacency matrix, 3, 8, 10, 12
adjusted relevance, **4**, 5, 19–28
algorithm
    HITS, **10**
    PageRank, **8**
    ProT, **12**

BHITS, 10

CLEVER, 10
coverage, **4**

eigenvalue, 8
    dominant, 3
eigenvector
    dominant, 8

HITS, **10**, 15, 16
    authority, **10**, 11, 14
    hub, **10**, 11, 14
    Randomised, 10, 16
    Subspace, 10
hyper-link, 3, 6

matrix
    adjacency, 3, 8, 10, 12
maximum spread, **4**, 5, 19–28
method
    power, 9

PageRank, **8**, 9, 14–16
    Topic-Sensitive, 9, 10, 14–16
power method, 8, 9
Propagation of Topic-relevance, **12**
ProT, **12**, 13–16

Randomised HITS, 10, 16
relevance, **4**
    adjusted, **4**, 5, 19–28

S^2ProT, 12–16
SALSA, 10
Spearman ρ, 5, 8, 14
Spearman Footrule Distance, 5, 8, 14
spread
    maximum, **4**, 5, 19–28
Subspace
    HITS, 10

Topic-Sensitive PageRank, **9**, 10, 14–16