Exploiting Community Behavior for Enhanced Link Analysis and Web Search

Dagstuhl Seminar 07071 "Web Information Retrieval and Linear Algebra Algorithms"
Technical Report extending [1]

Julia Luxenburger, Gerhard Weikum

Max-Planck Institute for Informatics, Database and Information Systems Group
Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany
{julialux|weikum}@mpi-inf.mpg.de

Abstract. Methods for Web link analysis and authority ranking such as PageRank are based on the assumption that a user endorses a Web page when creating a hyperlink to this page. There is a wealth of additional user-behavior information that could be considered for improving authority analysis, for example, the history of queries that a user community posed to a search engine over an extended time period, or observations about which query-result pages were clicked on and which ones were not clicked on after a user saw the summary snippets of the top-10 results. This paper enhances link analysis methods by incorporating additional user assessments based on query logs and click streams, including negative feedback when a query-result page does not satisfy the user demand or is even perceived as spam. Our methods use various novel forms of advanced Markov models whose states correspond to users and queries in addition to Web pages and whose links also reflect the relationships derived from query-result clicks, query refinements, and explicit ratings. Preliminary experiments are presented as a proof of concept.

Keywords. negative feedback, link analysis, web search, query logs

1 Introduction

Improving the ranking of web search results by means of link analysis and derived authority scores has become a de facto standard, with the PageRank [2] algorithm being the most prominent approach. However, the increasing amount of web spam and the continuous growth of low-quality web sites are major impediments to the viability of authority ranking in a world of exploding information and demanding users. On the other hand, the users' assessments of web pages should not be limited to the implicit endorsements by links. Rather, users can contribute in the form of explicit feedback, by marking search results as relevant, implicitly by clicking on search results, visiting certain pages (click streams), by blogs, wikis, and so forth. Moreover initiatives are arising towards a tagged web

Dagstuhl Seminar Proceedings 07071
Web Information Retrieval and Linear Algebra Algorithms
http://drops.dagstuhl.de/opus/volltexte/2007/1066

in which hyperlinks are no longer purely based on navigational purposes but augmented by semantic meaning, in its simplest form by "like" and "dislike" statements [3]. This calls for novel forms of extended authority analysis to harness the newly arising ways of assessments, especially expressions of disliking a page, which, to our knowledge, have not been addressed in the context of authority analysis.

PageRank completely ignores the different intentions that lead a web page author to create a hyperlink which may be purely navigational, or of recommending or disapproving flavor. The PageRank algorithm mimics a random surfer who starts on some page, then browses the web by following outgoing hyperlinks uniformly at random with probability ϵ, or re-starts by a random jump with probability $1-\epsilon$ (with uniformly selected jump target). This is formally modeled as a Markov chain, the unique equilibrium probability distribution of which yields stationary visiting probabilities, that consitute the vector of PageRank scores $\boldsymbol{p}$. Mathematically, PageRank is cast into the equation

$$\boldsymbol{p} = \epsilon \cdot \boldsymbol{r} + (1-\epsilon) \cdot A^T \boldsymbol{p}$$

where $\boldsymbol{r}$ denotes the random jump vector with $\sum_i \boldsymbol{r}_i = 1$, and A is the row-normalized adjacency matrix defined by the hyperlink structure of the web that already includes the treatment of dangling nodes.

Various approaches exist for how to exploit implicit feedback from query logs for web search. [4] employs query clustering for the identification of frequently asked questions. This method is, however, restricted to the very query context, and not able to take advantage of the gathered knowledge for an improvement of search result quality of previously unseen queries. [5] learns term correlations between terms occuring in clicked documents and terms constituting the corresponding queries for improved query expansion. [6] uses implicit feedback information of the current search session for better estimating query language models inside the KL-divergence retrieval model. [7] exploits query-log data to learn retrieval functions using a Support Vector Machine (SVM) approach. [8,9,3] point out the semantic difficulty of distrust propagation, but at the same time show the potential of considering negative endorsements. In the context of recommender systems, [9] aims at the prediction of pairwise trust of one node into another, however, they do not tackle the problem of absolute trust measures we address. [8] proposes facilitating PageRank-style distrust propagation by first computing PageRank on the trust relations and then subtracting the PageRank of sources of distrust statements from the ranks of their targets.

The approaches we propose build on our earlier work [10], based on a Markov-chain model with queries as additional nodes, additional edges that capture query refinements and result clicks, and corresponding transition probabilities. This prior work did, however, considers only positive feedback, inferred from a user clicking on a query result. The model could not express negative feedback from not clicking on a result although a lower-ranked result was clicked on. The methods of the current paper, on the other hand, support a much richer model that can handle also the case of non-clicked result pages, and moreover, can capture and exploit more general forms of negative assessment such as assigning trust levels to Web pages (e.g., marking a Web page as spam, low-quality, out-of-date, or untrusted [9]). For example, within a PageRank-style link analysis, if many users express distrust in a particular page then the authority (PageRank score mass) that this page receives from its in-link neighbors should be reduced. A key difficulty in exploiting both positive and negative assessment is that negative bias cannot be easily expressed in terms of probabilities, as probabilities are always non-negative and L1-normalized. We pursue several approaches that extend standard Markov models, one of which is based on a Markov reward model [11] where the assessment part is uncoupled from the random walk in the extended Web graph.

The rest of the paper is organized as follows. Section 2 introduces three different ways of integrating user assessments into Markovian authority propagation models. Preliminary experiments on two datasets are presented in Section 3.

2 Behavior-sensitive Authority

2.1 Data model

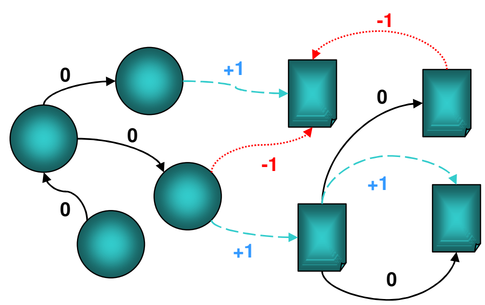

Fig. 1. Data model

As depicted in Figure 1, the graph model we consider is general enough to allow for typed nodes representing different entities, as well as tagged links carrying rating information. The displayed example graph indicates web pages by squared nodes, and queries by round ones. Directed links connecting them express categorical judgements , i.e., we distinguish the three ratings *positive* (+1), *neutral* (0), and *negative* (-1) - our models, however, can be easily extended to allow for more fine-grained quantifications. Thus we consider three link types depending on the rating associated with them. Let $\mathcal{E}$ denote the set of all links, $\mathcal{E}^+$ the set of links carrying a positive assessment, $\mathcal{E}^0$ the set of neutral, and $\mathcal{E}^-$ the set of negative links. Furthermore $\mathcal{S}$ is the set of all nodes with the subsets $\mathcal{S}^+$ and $\mathcal{S}^-$ denoting the sources of positive and negative links respectively.

2.2 QRank

QRank has been introduced in [10] to exploit implicit positive feedback obtained from query logs. QRank distinguishes between two node types, queries and web pages, and represents query-result clicks as well as query refinements by directed links from queries to pages and between queries respectively. To cast QRank into our general data model, we consider a variant of QRank that ignores virtual links based on textual similarity between documents and queries, repectively, and performs transitions uniformly at random. Random jumps are biased towards the sources of positive feedback whereby the bias strength is regulated by the parameter β.

The QRank model however faces some limitations in that it cannot model negative feedback. Assume a user marks a search result as irrelevant for a certain query. Given that some other user gave positive feedback on the very same relation, i.e., the QRank graph already contains a link from query A to document B, we can model the presence of negative feedback by reducing the transition probability from A to B with respect to all other links leaving A. In the case that there is no such link yet, we lack means to model negative feedback inside QRank. In the following we present a number of approaches that integrate negative endorsements.

2.3 QLoop*

The first algorithm, we propose, is not based on PageRank itself, but on a slight variant, the *self-loop algorithm*, which differs from PageRank only by the introduction of self-loops each node performs with probability δ, i.e,

$$\boldsymbol{s} = \epsilon \cdot \boldsymbol{r} + \delta \cdot \boldsymbol{s} + (1 - \epsilon - \delta) \cdot A^T \boldsymbol{s}$$

For the difference between the induced stationary visiting probabilities, $\boldsymbol{\pi}_p$ and $\boldsymbol{\pi}_s$, we can derive the following upper-bound in terms of the L_1-norm

$$||\boldsymbol{\pi_s} - \boldsymbol{\pi_p}||_1 \leq \frac{2 \cdot \delta}{\epsilon}$$

Proof. We have

$$\boldsymbol{\pi_p} = \epsilon \cdot \boldsymbol{r} + (1 - \epsilon) \cdot A^T \boldsymbol{\pi_p}$$

and

$$\boldsymbol{\pi_s} = \epsilon \cdot \boldsymbol{r} + \delta \cdot \boldsymbol{\pi_s} + (1 - \epsilon - \delta) \cdot A^T \boldsymbol{\pi_s}$$

and thus

$$\begin{aligned}
&\boldsymbol{\pi_s} - \boldsymbol{\pi_p} = \delta \cdot \boldsymbol{\pi_s} + (1 - \epsilon - \delta) \cdot A^T \boldsymbol{\pi_s} - (1 - \epsilon) \cdot A^T \boldsymbol{\pi_p} \\
\Leftrightarrow\ &\boldsymbol{\pi_s} - \boldsymbol{\pi_p} = \delta \cdot \boldsymbol{\pi_s} - \delta \cdot A^T \boldsymbol{\pi_s} + (1 - \epsilon) \cdot A^T (\boldsymbol{\pi_s} - \boldsymbol{\pi_p}) \\
\Leftrightarrow\ &\boldsymbol{\pi_s} - \boldsymbol{\pi_p} = (I - (1 - \epsilon) \cdot A^T)^{-1} (\delta \cdot \boldsymbol{\pi_s} - \delta \cdot A^T \boldsymbol{\pi_s})
\end{aligned}$$

As $I - (1-\epsilon) \cdot A^T$ is an M-matrix with $||I - (1-\epsilon) \cdot A^T||_1 = \epsilon$, we have $(I - (1-\epsilon) \cdot A^T)^{-1} \geq 0$ with L_1-norm $\frac{1}{\epsilon}$. Thus

$$
\begin{aligned}
||\boldsymbol{\pi_s} - \boldsymbol{\pi_p}||_1 &= ||(I - (1-\epsilon) \cdot A^T)^{-1}(\delta \cdot \boldsymbol{\pi_s} - \delta \cdot A^T \boldsymbol{\pi_s})||_1 \\
&\leq ||(I - (1-\epsilon) \cdot A^T)^{-1}||_1 \, ||\delta \cdot \boldsymbol{\pi_s} - \delta \cdot A^T \boldsymbol{\pi_s}||_1 \\
&\leq \frac{1}{\epsilon} \cdot ||\delta \cdot (I - A^T) \boldsymbol{\pi_s}||_1 \\
&\leq \frac{1}{\epsilon} \cdot \delta \cdot ||I - A^T||_1
\end{aligned}
$$

Let $K = I - A^T$. Then the absolute column sum of the jth column of K is

$$\sum_i |k_{ij}| = 1 - a_{jj} + \sum_{i \neq j} |-a_{ij}| = 1 - a_{jj} + 1 - a_{jj} = 2 - 2 \cdot a_{jj}$$

as A^T is column-stochastic. Thus

$$||K||_1 = \max_j \{2 - 2 \cdot a_{jj}\} = 2$$

in case a nondangling node exists. Thus $||\boldsymbol{\pi_s} - \boldsymbol{\pi_p}||_1 \leq \frac{2 \cdot \delta}{\epsilon}$.

The change in ranking order is however more important for web search than changes in terms of absolute ranking scores. Just by reasoning on the defining equation of the self-loop algorithm which turns into PageRank as $\delta \to 0$, we find that authority scores under both algorithms share some base contribution which stems from random jumps, and differ in how much authority is propagated via incoming links. Thus self-loops reduce the influence of predecessors in favor of some selfishness of always keeping a fraction of own authority. As a consequence low-indegree nodes experience a slight boost in score with self-loops being added, while auhoritative pages under PageRank undergo small perturbation due to the reduced authority that is propagated to them and recursive changes in authority propagation. This intuition is experimentally underpinned by comparing scores of corresponding nodes under the two algorithms (Figure 2 plots nodes and their logarithmic-scaled scores in descending order of PageRank scores).

Analogously, we may consider a self-loop augmented variant of QRank, coined *QLoop*, forming the basis of a holistic approach to integrate both positive and negative endorsements into link analysis. To infer a notion of community-level authority, we consider a hybrid method, *QLoop**, that models positive ratings the way QLoop does, and translates negative ratings into node-specific loop and jump probabilities. To make successors of a punished node not benefit from changes in the self-loop probability δ, we re-distribute the remaining probability mass by increasing the respective random jump probability. The amount by which we decrease the self-loop probability of a negatively judged node depends on the authority scores of its predecessors which are estimated by computing

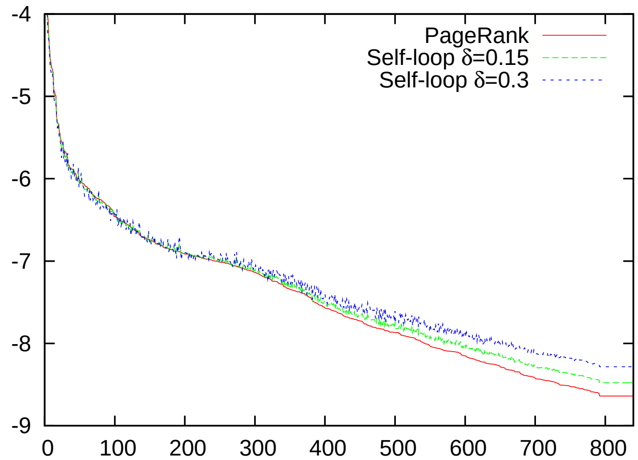

Fig. 2. PageRank vs Self-loop algorithm

QRank in a pre-processing step. That way we facilitate an intertwining of assessment and authority propagation. To back our intuition that a decrease of the self-loop probability δ^* of a selected node i^* indeed results in a decreased score, we reason on the defining equation of the self-loop algorithm. Making the contributions of incoming links and dangling pages encoded in the A matrix explicit and assuming i^* is non-dangling, we have

$$\boldsymbol{s}_{i^*} = \epsilon \cdot \boldsymbol{r}_{i^*} + \delta^* \cdot \boldsymbol{s}_{i^*} + (1 - \epsilon - \delta) \cdot \left[\sum_{(j,i^*) \in \mathcal{E}} \frac{\boldsymbol{s}_j}{o_j} + \sum_{j \in DP} \frac{\boldsymbol{s}_j}{|\mathcal{S}|} \right]$$

where o_j denotes the outdegree of j and DP is the set of dangling pages. Thus under the assumption that no other node undergoes changes in ϵ and δ, $\delta^* < \delta$ implies a reduced score for i^*.

Definition 1 (QLoop*). *Let $\boldsymbol{\pi}_q$ denote the stationary visiting probabilities under QRank, and A the adjacency matrix over $\mathcal{E}^+ \cup \mathcal{E}^0$ including the handling of dangling nodes. Then QLoop* scores, denoted by $\boldsymbol{q}_{loop*}$, are defined as follows*

$$\boldsymbol{q}_{loop*} = \epsilon \cdot \boldsymbol{r} + \delta \cdot W^T \boldsymbol{q}_{loop*} + (1 - \epsilon - \delta) \cdot A^T \boldsymbol{q}_{loop*}$$

with random jumps being biased according to

$$\boldsymbol{r}_i = \begin{cases} \frac{\beta}{|\mathcal{S}^+ \cup \mathcal{S}^-|} & , \quad \textit{if } i \in \mathcal{S}^+ \cup \mathcal{S}^- \\ \frac{1-\beta}{|S - (\mathcal{S}^+ \cup \mathcal{S}^-)|} & , \quad \textit{otherwise} \end{cases}$$

and self-loops adjusted according to

Case 1: $\exists k \in \mathcal{S}^- : (k, i) \in E^-$

1. ***with normalization***

$$w_{ij} = \begin{cases} 1 - \sum\limits_{\{k|(k,i)\in E^-\}} \frac{\boldsymbol{\pi}_q(k)}{|\{l\in\mathcal{S}|(k,l)\in E^-\}|},\ \mathit{if}\ i = j \\ \frac{1}{|\mathcal{S}|-1} \cdot \sum\limits_{\{k|(k,i)\in E^-\}} \frac{\boldsymbol{\pi}_q(k)}{|\{l\in\mathcal{S}|(k,l)\in E^-\}|},\ \mathit{if}\ i \neq j \end{cases}$$

2. ***without normalization***

$$w_{ij} = \begin{cases} 1 - \sum\limits_{\{k|(k,i)\in E^-\}} \boldsymbol{\pi}_q(k),\ \mathit{if}\ i = j \\ \frac{1}{|\mathcal{S}|-1} \cdot \sum\limits_{\{k|(k,i)\in E^-\}} \boldsymbol{\pi}_q(k),\ \mathit{if}\ i \neq j \end{cases}$$

Case 2: $\nexists k \in \mathcal{S}^- : (k,i) \in E^-$

$$w_{ij} = \begin{cases} 1 & , \quad \mathit{if}\ i = j \\ 0 & , \quad \mathit{if}\ i \neq j \end{cases}$$

Theorem 1. *QLoop* defines an ergodic Markov chain.*

Proof. It can be shown easily that W is a stochastic matrix which implies that QLoop* defines a Markov chain. Irreducibility is ensured by random jumps, and aperiodicity is a consequence of the self-loops. From Markov chain theory, we know that a finite, irreducible and aperiodic Markov chain, is also ergodic. Thus QLoop* converges.

2.4 Behavior-sensitive Jumps

In resemblance to personalized PageRank [12], we propose to integrate additional assessments into the process of authority propagation by the aggregation of endorsements and disapprovals into a biased random jump vector. Thus nodes receiving positive ratings are more often starting states of a new path the random surfer pursues than nodes judged to be of poor quality. Let $\hat{R}$ be a matrix of rewards such that

$$\hat{r}_{ij} = \begin{cases} -1, \text{ if } (i,j) \in \mathcal{E}^- \\ 0, \text{ if } (i,j) \in \mathcal{E}^0 \\ 1, \text{ if } (i,j) \in \mathcal{E}^+ \end{cases}$$

Depending on how we choose to aggregate recommendations and disfavors, we distinguish between three incarnations of *behavior-sensitive* random jump vectors, denoted by $\boldsymbol{r}_{BS}$ in the following.

Uniform: $\boldsymbol{r}_{BS}(j) = \sum_i \hat{r}_{ij}$
Normalized: $\boldsymbol{r}_{BS}(j) = \sum_i \frac{\hat{r}_{ij}}{\sum_j |\hat{r}_{ij}|}$
Weighted: $\boldsymbol{r}_{BS}(j) = \sum_i \hat{r}_{ij} \cdot \boldsymbol{\pi}(i)$

This first aggregation of ratings is followed by a normalization step, the addition of the one vector, and a final re-normalization step yielding the final jump vectors. In the weighted feedback aggregation scenario, stationary visiting probabilities under PageRank ($\mathcal{E}^0$ defines the link structure) with β-biased random jumps to nodes in $\mathcal{S}^- \cup \mathcal{S}^+$ serve as authority scores $\boldsymbol{\pi}$. The following theorem gives an upper bound in terms of L_1-norm on the difference between the steady-state probability distributions of PageRank and behavior-sensitive random jumps.

Theorem 2. *Let $\boldsymbol{\pi}_{BS}$ denote the unique equilibrium probability distribution under behavior-sensitive random jumps. Then $||\boldsymbol{\pi}_{BS} - \boldsymbol{\pi}_p||_1 \leq ||\boldsymbol{r}_{BS} - \boldsymbol{r}||_1$.*

Proof.

$$\begin{aligned} \boldsymbol{\pi}_{BS} - \boldsymbol{\pi}_p &= \epsilon \cdot (\boldsymbol{r}_{BS} - \boldsymbol{r}) + (1-\epsilon) \cdot A^T (\boldsymbol{\pi}_{BS} - \boldsymbol{\pi}_p) \\ \Leftrightarrow \boldsymbol{\pi}_{BS} - \boldsymbol{\pi}_p &= (I - (1-\epsilon) \cdot A^T)^{-1} \cdot \epsilon \cdot (\boldsymbol{r}_{BS} - \boldsymbol{r}) \end{aligned}$$

Thus

$$\begin{aligned} ||\boldsymbol{\pi}_{BS} - \boldsymbol{\pi}_p||_1 &\leq \epsilon \cdot ||(I - (1-\epsilon) \cdot A^T)^{-1}||_1 \cdot ||\boldsymbol{r}_{BS} - \boldsymbol{r}||_1 \\ &\leq \epsilon \cdot \frac{1}{\epsilon} \cdot ||\boldsymbol{r}_{BS} - \boldsymbol{r}||_1 \\ &\leq ||\boldsymbol{r}_{BS} - \boldsymbol{r}||_1 \end{aligned}$$

2.5 Markov Reward Model

Inspired by the use of Markov reward models in the field of performance and dependability analysis, we propose to augment the graph model representing the hyperlink structure of the web with an additional reward structure. Thereby each page is associated with a reward accumulator variable - collectively denoted by the vector $\boldsymbol{g}$, which is updated each time the page is visited depending on the transition's reward. This reward depends on the transition's source and target and is derived from the query-log and click-stream information as well as explicit page assessments. With $\hat{R} = (\hat{r}_{ij})$ denoting a reward matrix as defined in Section 2.4, each transition along $(i,j) \in \mathcal{E}$ results in an update of the vector $\boldsymbol{g}$ according to $\boldsymbol{g}_n(j) = \boldsymbol{g}_{n-1}(j) + \hat{r}_{ij}$. Then the long-run average reward each node accumulates,

$$\boldsymbol{g}_\infty(i) = \lim_{n\to\infty} \frac{1}{n} \cdot \boldsymbol{g}_n(i)$$

gives an assessment-based measure of its quality. Thereby the contribution of each rating is implicitly weighted by the authority of its source given by how often it is visited during a random walk. In the following we present a theorem that allows us to compute the long-run average reward of each node efficiently.

Theorem 3. *Let $A = (a_{ij})$ denote a transition probability matrix defining a Markov chain, and $\boldsymbol{\pi}$ be the corresponding stationary visiting probability distribution. Then*

$$\lim_{n\to\infty} \frac{1}{n} \cdot \boldsymbol{g}_n(i) = \lim_{n\to\infty} \frac{1}{n} \cdot \sum_{k=1}^{n} \hat{r}_{s_k s_{k+1}} = \sum_{(j,i)\in E} \hat{r}_{ji} \cdot a_{ji} \cdot \boldsymbol{\pi}_j$$

Proof. This proof follows the lines of a related proof found in [11] and can be similarly found in [13]. Assume A defines an irreducible, time-homogeneous and aperiodic Markov chain. Then state i is reachable from every other state s $\in \mathcal{S}$ with probability $f_{si} = 1$. The mean recurrence time μ_{ii} of returning to state i is by the same argument $< \infty$. The Markov chain can be seen as a regenerative process with epochs at which the process visits state i as regeneration epochs. Now consider the long-run average reward per time unit

$$\lim_{n\to\infty} \frac{1}{n} \cdot \sum_{k=1}^{n} \hat{r}_{s_k s_{k+1}}$$

First assume that state i is the initially visited state. Then define a cycle between the two successive visits of state i. The expected cycle length equals the mean recurrence time μ_{ii} and is finite. By the renewal-reward theorem it holds

$$\lim_{n\to\infty} \frac{1}{n} \cdot \sum_{k=1}^{n} \hat{r}_{s_k s_{k+1}} = \frac{E(\text{reward earned during one cycle})}{E(\text{length of one cycle})}$$

with probability 1. The reward earned during one cycle is $0 + r(\text{transition from predecessor j to i})$

With

$$E(\text{number of visits to j between two successive visits to i}) = \frac{\pi_j}{\pi_i}$$

we obtain

$$E(\text{reward earned during one cycle}) = \sum_{(j,i)\in E} \hat{r}_{ji} \cdot a_{ji} \cdot \frac{\pi_j}{\pi_i}$$

Since $E(\text{length of one cycle}) = \mu_{ii} = \frac{1}{\pi_i}$, it holds

$$\lim_{n\to\infty} \frac{1}{n} \cdot \sum_{k=1}^{n} \hat{r}_{s_k s_{k+1}} = \sum_{(j,i)\in E} \hat{r}_{ji} \cdot a_{ji} \cdot \pi_j$$

Consider now the Markov process that starts in any arbitrary state $s \neq i$. Since $\forall s \in \mathcal{S} : f_{si} = 1$ the process will eventually reach state i. Let t be the time step of entering state i. Then

$$\begin{aligned} &\frac{1}{n} \cdot \sum_{k=1}^{n} \hat{r}_{s_k s_{k+1}} \\ &= \frac{1}{n} \cdot \sum_{k=1}^{t} \hat{r}_{s_k s_{k+1}} + \frac{1}{n} \cdot \sum_{k=t+1}^{n} \hat{r}_{s_k s_{k+1}} \end{aligned}$$

Letting $n \to \infty$, the first term on the right-hand side of the equation tends to zero, whereas the second term as shown before converges to $\sum_{(j,i)\in E} \hat{r}_{ji} \cdot a_{ji} \cdot \pi_j$.

QRank computed on the total set of edges $\mathcal{E}$ serves as our baseline for the derivation of transition probabilities (a_{ij}) and stationary visiting probabilities $\boldsymbol{\pi}$. We compute the final ranking scores $\boldsymbol{\pi}_g$, coined *QReward*, as a linear combination of the re-normalized long-run average reward with the underlying authority scores as follows.

$$\boldsymbol{\pi}_g = \alpha \cdot \boldsymbol{g}_\infty + (1-\alpha) \cdot \boldsymbol{\pi}$$

QDiscounter In addition, we consider a slight variant of QReward, coined *QDiscounter*, that is lazy in computing the long-run average rewards and simply omits the multiplication with the transition probabilities, i.e.,

$$\lim_{n\to\infty} \frac{1}{n} \cdot \boldsymbol{g}_n(i) \approx \sum_{(j,i)\in E} \hat{r}_{ji} \cdot \boldsymbol{\pi}_j$$

The underlying visiting probabilities $\boldsymbol{\pi}$ are computed using QRank on $\mathcal{E}^0$ and β-biasing the nodes in $\mathcal{S}^- \cup \mathcal{S}^+$.

Table 1 summarizes the various ranking methods we consider, and indicates for each method the link structure and random jumps it builds on, as well as the parameter it requires.

Algorithm	**Links**	**Parameter**	**Random Jump**
PageRank	$\mathcal{E}^0$	ϵ	Uniform
QRank	$\mathcal{E}^0 \cup \mathcal{E}^+$	ϵ, β	Biased ($\mathcal{S}^+$)
QLoop*	$\mathcal{E}^0 \cup \mathcal{E}^+$	ϵ, β, δ	Biased ($\mathcal{S}^+ \cup \mathcal{S}^-$)
BS Jumps	$\mathcal{E}^0$	ϵ, β	Biased
QReward	$\mathcal{E}$	ϵ, β, δ, α	Uniform
QDiscounter	$\mathcal{E}^0$	ϵ, β, δ, α	Uniform

Table 1. Overview of Authority Ranking Methods

3 Preliminary Experiments

3.1 Data collection

As datasets with positive and negative endorsements are difficult to obtain outside the commercial search engine companies, we created our own data collections based on two datasets with very different properties: an excerpt of the web pages of the Wikipedia Encyclopedia, and a linkage graph constituted by product data of `Amazon.com`.

Wikipedia Starting from overview pages about geography, history, film, and music we crawled 72482 documents on a downloaded dump of Wikipedia to build a thematically concentrated dataset and indexed it by our own prototype search engine. For query session generation, we asked 18 volunteers, students with diverse backgrounds (law, psychology, intercultural communication, etc) to search our data collection. We provided some creativity help in the form of Trivial Pursuit questions and asked the volunteers to concentrate on the categories geography, history, and entertainment. But they were still allowed to freely choose their queries or follow some personal interests to simulate real web search. Parsing the generated browser history files, we obtained 542 queries, 760 query result clicks (implicit positive feedback), 290 query refinements and 1987 implicit negative feedback links. We interpreted each non-clicked document appearing above a clicked one as negative feedback, driven by the justification that the user saw the summary snippets of these pages and intentionally skipped them.

Query	PageRank	QRank $\beta = 1.0$	QLoop* $\beta = 1.0$ $\delta = 0.3$	QReward $\beta = 1.0$ $\alpha = 0.2$	QDiscounter $\alpha = 0.8$	Uniform BS-Jump $\epsilon = 0.5$
birthplace mozart	0.51	0.65	0.68	0.64	0.64	0.53
brazil cities	0.37	0.43	0.4	0.48	0.39	0.39
political system of china	0.35	0.57	0.65	0.53	0.42	0.34
free elections german democratic republic	0.13	0.18	0.2	0.16	0.33	0.14
Egypt pyramids	0.2	0.55	0.56	0.55	0.65	0.25
Napoleon exile	0.82	0.7	0.72	0.79	0.49	0.85
Harrison Ford movie	0.16	0.41	0.41	0.32	0.35	0.19
French wine	0.83	0.39	0.37	0.39	0.33	1
John Paul II	0.25	0.5	0.5	0.5	0.5	0.25
official language Singapore	0.27	0.29	0.29	0.33	0.5	0.33
last play by Shakespeare	0.8	0.81	0.8	0.8	0.79	0.81
Nelson Mandela prison	0.67	0.71	0.78	0.71	0.69	0.66
firefighter New York	0.25	0.23	0.25	0.18	0.45	0.26
constitutional supreme court	0.73	0.73	0.73	0.8	0.81	0.63

Table 2. MAPs of evaluation queries on Wikipedia

For evaluation, we chose 14 queries (see Table 2) at random from a set of queries that had been posed by users during query session generation and for which textual-similarity based retrieval yielded result sets of size at least 50. 10 out of the 14 queries were also associated with negative assessments.

Amazon With the help of the Amazon E-Commerce web service we constructed a graph similar in structure to the enhanced web graph we obtained from the Wikipedia data. We distinguish two node types, items and customers, with the latter corresponding conceptually to the previous notion of queries. We establish a link from item A to item B whenever B is said to be similar to A, i.e., customers who bought A also bought B. Furthermore a customer reviewing a particular item is represented by a link which is associated with a positive reward for a rating greater than three stars, and a negative reward for ratings of less than three stars. Ratings of exactly three stars result in neutral links as well. In that manner we constructed a graph of 247688 items, 607663 customers, 1258487 neutral, 912775 positive and 138813 negative reward links.

3.2 Methodoloy

Query result rankings are derived as follows. For each query, we construct a seed set consisting of the top-50 query results solely based on textual similarity. For the Wikipedia dataset these are retrieved according to Okapi BM25 [14], whereas Amazon builds on textual similarity scores on the editorial reviews of products, with scores computed by the Oracle Text product (which we used as a backend in our implementation). The query results are then re-ordered according to our pre-computed ranking schemes, and in case of ties we fall back to the text-based scoring. As quality assessments are usually sparse, we vary the graphs on which rankings are to be computed to strengthen the influence of quality judgements by means of back-links. These are neutral reversed links of rating-carrying links in $\mathcal{E}^- \cup \mathcal{E}^+$. That way we improve the reachability of nodes in $\mathcal{S}^+ \cup \mathcal{S}^-$ which are with our current datasets often solely reachable via random jumps.

3.3 Results

For evaluation of search result quality, we computed the top-15 result rankings on Wikipedia, presented 8 volunteers an unordered list of URLs occuring in at least one result ranking for the given query and asked them to mark the relevant ones (possibly after consulting the linked result page). We had each query evaluated by 3 different users and took their majority vote as the final relevance assessment. That way the obtained relevance assessments are consistent over all evaluated rankings. To account for the ranks at which relevant documents occur in a ranking, we chose to compute the mean average precision (MAP) of each query that is sensitive to re-orderings in the result set, and defined as

$$\sum_{r=1}^{15} \frac{precision@r * rel(r)}{\#\text{relevant docs}}$$

where r denotes the rank, and rel(r) indicates whether the document at rank r is a relevant one.

Table 2 depicts the resulting MAP values for each query evaluated on the Wikipedia dataset and some representative ranking schemes. The averaged MAP values as well as the standard deviation across queries for each considered method are depicted in Table 3. QLoop* achieves improvements over both PageRank and QRank regardless of the values we chose for β and δ. The MAP values of the normalized variant of QLoop* coincide with those of the non-normalized version, indicating that normalization plays a minor role for ranking. The family of behavior-sensitive random jumps outperforms PageRank, but does not reach the performance of QRank. Coding ratings inside the random jump vector seems to have little effect, even under extreme parameter settings. The Markov reward model and its approximation are the most promising approaches with QDiscounter yielding significant gains in MAP compared to all other methods. QDiscounter achieved MAP values around 55 percent across the spectrum of choices for α, compared to 51 percent for QRank and merely 45 percent for standard PageRank. Interestingly, QDiscounter did not benefit from the introduction of back-links, but showed better results with the normal graph structure. This remains to be further investigated on different datasets.

Table 4 shows top-5 result rankings (titles of Wikipedia pages) for the query "Political system of China". To better understand the effects observed, Table 3.3 lists the documents in the top-50 result set based on textual similarity that received positive or negative long-run average rewards due to the implicit feedback obtained from query-logs. When comparing these two tables we see the different extents to which the proposed methods combine endorsements with standard link analysis.

Table 6 shows top-5 rankings of books computed on the Amazon dataset for the Google Zeitgeist query "mountain bike". In contrast to the Wikipedia dataset where query-log data was sparse, Amazon offers a larger amount of rating data, and thus a more balanced ratio of customer to item nodes. Again we observe a varying strength with which ratings are incorporated ranging from the behavior-sensitive jumps which are closest to PageRank, over QRank and QLoop* to the Markov reward model approaches which show the most significant changes. The way our behavior-sensitive approaches favor specialized books on mountain bikes over "traditional" authorities on the subject of traveling, like the "Fodor's" series, shows their effectiveness.

4 Conclusion

We presented three novel algorithmic frameworks to incorporate additional user assessments into web link analysis, and underpinned their potential by preliminary experiments on two datasets. Currently, we are acquiring larger datasets with a broad spectrum of user assessments to further investigate the proposed algorithms.

	PageRank	**QRank $\beta = 1.0$**	**QRank $\beta = 0.5$**
MAP	0.4511	0.5099	0.5136
Deviation	0.09	0.08	0.08

	QLoop*					
β	1.0			0.5		
δ	0.15	0.3	0.4	0.15	0.3	0.4
MAP	0.52	0.53	0.53	0.52	0.52	0.52
Deviation	0.01	0.08	0.06	0.03	0.03	0.01

	Normalized QLoop*					
β	1.0			0.5		
δ	0.15	0.3	0.4	0.15	0.3	0.4
MAP	0.52	0.53	0.53	0.52	0.52	0.52
Deviation	0.08	0.03	0.04	0.08	0.08	0.04

	BS-Jump							
	weighted				**uniform**		**normalized**	
β	1.0	1.0	0.5	0.5	-			
ϵ	0.25	0.5	0.25	0.5	0.25	0.5	0.25	0.5
MAP	0.46	0.47	0.46	0.47	0.46	0.47	0.46	0.47
Deviation	0.06	0.04	0.06	0.06	0.03	0.09	0.03	0.06

Markov reward model					
α	0.2	0.4	0.5	0.6	0.8
QReward $\beta = 0.5$					
MAP	0.49	0.48	0.48	0.47	0.51
Deviation	0.002	0.02	0.09	0.09	0.002
QReward $\beta = 1.0$					
MAP	0.51	0.49	0.49	0.48	0.52
Deviation	0.1	0.0001	0.05	0.03	0.04
QDiscounter $\beta = 0.5$					
MAP	0.49	0.49	0.52	0.52	0.52
Deviation	0.05	0.002	0.005	0.07	0.007
QDiscounter without back-links $\beta = 0.5$					
MAP	0.53	0.53	0.54	0.55	0.56
Deviation	0.0004	0.08	0.005	0.03	0.04

Table 3. Average MAP/std. deviation on Wikipedia

References

1. Luxenburger, J., Weikum, G.: Exploiting community behavior for enhanced link analysis and web search. In: WebDB. (2006)
2. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. Technical report, Stanford (1998)
3. Massa, P., Hayes, C.: Page-rerank: using trusted links to re-rank authority. Technical report, ITC - IRST (2005)
4. Wen, J.R., Nie, J.Y., Zhang, H.J.: Query clustering using user logs. In: TIS. (2002)
5. Cui, H., Wen, J.R., Nie, J.Y., Ma, W.Y.: Query expansion by mining user logs. In: TKDE. (2003)
6. Shen, X., Tan, B., Zhai, C.: Context-sensitive information retrieval using implicit feedback. In: SIGIR. (2005)
7. Joachims, T.: Optimizing search engines using clickthrough data. In: KDD. (2002)
8. Guha, R.: Open rating systems. Technical report, Stanford (2003)
9. Guha, R., Kumar, R., Raghavan, P., Tomkins, A.: Propagation of trust and distrust. In: WWW. (2004)
10. Luxenburger, J., Weikum, G.: Query-log based authority analysis for web information search. In: WISE. (2004)
11. Tijms, H.C.: A First Course in Stochastic Models. John Wiley & Sons (2003)
12. H.Haveliwala, T.: Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search. In: TKDE. (2003)
13. Wolff, R.W.: Stochastic Modeling and the Theory of Queues. Prentice Hall (1989)
14. Robertson, S.E., Walker, S.: Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In: SIGIR. (1994)

PageRank	**QRank** $\beta = \mathbf{1}$
China	China
People's Republic of China	One country - two systems
List of countries	People's Republic of China
County	Communist state
Chinese language	List of countries
QLoop$^{*}\beta = \mathbf{1}, \delta = \mathbf{0.3}$	**QReward** $\beta = \mathbf{1}, \alpha = \mathbf{0.2}$
One country - two systems	One country - two systems
China	China
People's Republic of China	Prison
Communist state	List of countries
List of countries	Communist state
QDiscounter	
$\alpha = \mathbf{0.4}$	$\alpha = \mathbf{0.8}$
China	Prison
Prison	Communist state
Communist state	Party discipline
Party discipline	One country - two systems
One country, two systems	China
BS-Jump uniform	
$\epsilon = \mathbf{0.25}$	$\epsilon = \mathbf{0.5}$
China	China
People's Republic of China	County
County	People's Republic of China
List of countries	Hong Kong
Chinese language	Chinese language

Table 4. Top-5 for "Political System of China"

Positive	One country - two systems, Prison, Communist state, Party discipline
Negative	People's Republic of China, China, Vice President, Chinese language, Mandarin linguistics, Clash of Civilizations, Galileo positioning system

Table 5. Positively/negatively rewarded docs of "Political System of China"

PageRank
Fodor's Prague and Budapest
Bobke II
Mountain Biking Colorado
Zinn and the Art of Mountain Bike Maintenance
Bicycling Magazine's Complete Guide to Bicycle Maintenance and Repair for Road and Mountain Bikes
QRank $\beta = \mathbf{1}$
Zinn and the Art of Mountain Bike Maintenance
Bobke II
Bicycling Magazine's Complete Book of Road Cycling Skills
Bicycling Magazine's Complete Guide to Bicycle Maintenance and Repair for Road and Mountain Bikes
Fodor's Prague and Budapest
QLoop$^*\beta = \mathbf{1}, \delta = \mathbf{0.3}$
Zinn and the Art of Mountain Bike Maintenance
Bobke II
Bicycling Magazine's Complete Book of Road Cycling Skills
Bicycling Magazine's Complete Guide to Bicycle Maintenance and Repair for Road and Mountain Bikes
Fodor's Japan
QReward $\beta = \mathbf{1}, \alpha = \mathbf{0.8}$
Zinn and the Art of Mountain Bike Maintenance
Bobke II
Bicycling Magazine's Complete Book of Road Cycling Skills
Bicycling Magazine's Complete Guide to Bicycle Maintenance and Repair for Road and Mountain Bikes
Exploring the Black Hills and Badlands: A Guide for Hikers, Cross-Country Skiers, & Mountain Bikers
QDiscounter $\alpha = \mathbf{0.4}$
Mountain Bike! Southern Utah: A Guide to the Classic Trails
Fodor's Prague and Budapest
Bobke II
Zinn and the Art of Mountain Bike Maintenance
Mountain Biking Colorado
BS-Jump uniform $\epsilon = \mathbf{0.25}$
Fodor's Prague and Budapest
Bobke II
Zinn and the Art of Mountain Bike Maintenance
Bicycling Magazine's Complete Guide to Bicycle Maintenance and Repair for Road and Mountain Bikes
Mountain Biking Colorado

Table 6. Top-5 rankings for "mountain bike" on Amazon