Page-reRank: using trusted links to re-rank authority

Paolo Massa   Conor Hayes
ITC/iRst
Via Sommarive 14 - I-38050 Povo (TN) - Italy
massa, hayes@itc.it

Abstract

The basis of much of the intelligence on the Web is the hyperlink structure which represents an organising principle based on the human facility to be able to discriminate between relevant and irrelevant material. Second generation search engines like Google make use of this structure to infer the authority of particular web pages. However, the linking mechanism provided by HTML does not allow the author to express different types of links such as positive or negative endorsements of page content. Consequently, algorithms like PageRank produce rankings that do not capture the different intentions of web authors. In this paper, we review some of the initiatives for adding simple semantic extensions to the link mechanism. Using a large real world data set, we demonstrate the different page rankings produced by considering extra semantic information in page links. We conclude that Web intelligence would benefit in adoption of languages that allow authors easily encode simple semantic extensions to their hyperlinks.

1. Introduction

The idea of Web intelligence has been inextricably linked with the often described problem of information overload. Early research on Web intelligence viewed the web as a large, unstructured, distributed data base where the goal was to index resources so that users could easily find them. As such, early search engines considered the topology of the web to be flat, where the goal was to crawl as much content as possible. However, the flat approach wouldn't allow search engines to discriminate between documents of varying quality, authority or relevance. An answer to this problem emerged when researchers began to consider the hyperlink structure of the web as a means of inferring page relevance. Whereas early search engines used links simply as means to find additional content to index, more recent algorithms, for example PageRank [10], weighted pages according to how many other pages linked to them. The key observation is that the hyperlink structure of the web represents an organising principle based on

the human facility to be able to discriminate between relevant and irrelevant material. People who create web pages hand code links to other pages after judging the content on those pages. Thus, links explicitly characterize a filtering of content by human intelligence. By tapping this intelligence, *Google.com* returns pages that not only match the submitted query but have been linked to many other pages in the web. As many of the current research initiatives on Web intelligence are based on mining the link structure of the web we consider the semantics of using hyperlinks in our web documents. We point out that a link may have several meanings: referential, endorsing or criticising. The unary nature of the link mechanism in the Hypertext Markup Language (HTML) means that these meanings cannot be communicated. For example, the PageRank algorithm considers all links into a page to be a vote for that page. However, this is not always the case and Google bombers and blog spammers have exploited this weakness in order to boost the PageRank score of their sites.

Consequently, we consider the semantics required of a new linking structure and examine several existing proposals for a richer linking structure for the Web. We argue that any change to the hyperlink facility must be easily understandable by the ordinary users of the Web. In order to test the effect of a more expressive linking structure, we examine the different rankings produced by PageRank for a large real world data set where links are encoded using positive or negative trust statements. We see that the current unary link structure only allows PageRank to indicate how much attention is devoted to a particular page rather than how positively endorsed it is. We then consider what it means to have a graph based on negative links.

The paper is structured as follows. In Section 2 we give an overview of the hyperlink facility and describe the problems associated with it. In Section 3 we introduce and discuss various proposals to give web authors a more expressive linking language. In Section 4 we introduce the *Epinions.com* data set and we describe our experiments on ranking using the PageRank algorithm. We discuss our results in Section 5.

2. Linking on the Web

The term "hyperlink" was defined in 1965 by Ted Nelson during the Project Xanadu. In several books and articles published between 1964 and 1980, Nelson described his conception of a network of documents through which the reader makes her own trail by linking documents together. However, a research group led by Douglas Engelbart developed the *oNLine System* (NLS), the fist hyperlink system for connecting separate documents[1]. The prototype of the WWW was the Enquire programme developed by Tim Berners-Lee at CERN in 1980 in order to allow him to store and retrieve information about the structure of a system, such as the relationships and dependencies between people, programs, machines and ideas. The linking structure of Enquire allowed the author to define the type of relationship between two documents: such as part-of, made-by, similar-to. To extend the Enquire initiative to work in a shared information space, Berners-Lee developed the HTTP protocol and the HTML language for the World Wide Web. A crucial factor of the HTML language was its simplicity - a factor which allowed its rapid uptake. HTML contains only structural markup. The difficult task of agreeing a semantic mark-up language for the web was deferred. Likewise, the standard link semantics of the Enquire programme were discarded in favour of unary links that we have today on the Web.

In HTML, the anchor element ⟨a⟩ creates a link. The href attribute specifies the URL of the page resource being linked to. The description of the link known as anchor text is given between the opening and closing anchor elements.

⟨a href=″URL″ title=″a hover box text″⟩link description⟨/a⟩

Although, there are no predefined semantics for links in HTML, web authors generally use the anchor text to briefly describe the resource being linked to. Search engines attach a lot of weight to terms used in anchor text because they can represent a summary of the web author's impression or view of the web site. For example, the *Google* search engine will rank a page higher if all the sites that link to it use the same terms in their anchor text. This has led to a phenomenon known as *Google* Bombing whereby inappropriate anchor text is agreed in advance by group of web authors linking to a site of a person or organisation. A query to *Google* using the anchor text will return the web page of the victim even though the page itself does not contain the query terms[2].

So although the author can provide a free text description of the purpose of the link, machines are not sophisticated enough to detect when the author is being truthful, playful or malicious. This is problematic because the second generation of search engines such as *Google*, rely upon the link structure of the web to infer quality and relevance of search results. In the next section we look at some proposals for providing machine readable semantics to hyperlinks and we discuss the requirements and implications of such proposals.

3. Current improvements proposals

The second generation search engines have been able to improve search performance by considering and exploiting the hyperlink structure of the Web. PageRank [10], one of the algorithms powering *Google.com*, was the first with notable success in doing this. The basic idea of PageRank is that if page A links to page B, then the creator of page A is implicitly assigning some importance to page B. The PageRank or importance of page A, in turn, depends on the number of pages linking to it and on their importance. Intuitively, if many important pages point to a certain page, that page should be important as well. The notion of importance in PageRank is based simply on a measure of its 'connectedness' within the graph of the World Wide Web. Any semantics attached to the links by the author are not considered. An intuitive justification of PageRank is that it is the probability that a surfer starting at a random page and choosing a links at random will visit the page [3]. The randomness of the surfer's behaviour illustrates the insensitivity of the algorithm to the intentions of web authors. Thus a high PageRank score does not necessarily always correspond to positive endorsement. Authors can link to a page because they disapprove of the content expressed there and they may want to comment upon it: for example, linking to the web site of a disliked political candidate or a site with offensive or controversial content. In this cases, a link may simply be referential or constitute a negative endorsement.

It seems that humans tend to speak more about what they like and less about what they don't like (we provide some evidence of it in Section 4). For this reason, the adopted heuristics works most of the time. However, as web authors are often aware of increasing the PageRank of a page they link to, there is an incentive not to link to pages they *disagree* with. In this sense, the limited expressiveness of the language may be affecting the way authors encode their beliefs and opinions. We would argue that, in order to foster discussion and creativity, it is necessary to express both positive and negative preferences as well as the reasons for such preferences.

At the moment, it is not possible for an author to express a machine readable "semantics" for the link. In the following we will review some current initiatives to express "semantics" with links.

VoteLinks [12] is a microformat proposed by *Technorati.com* in which they propose a set of three new values for the rel attribute of the a tag in HTML. The values are vote-for, vote-abstain and vote-against and represent agree-

---

[1] http://sloan.stanford.edu/mousesite/1968Demo.html
[2] A list of successful Google Bombs is given here: http://en.wikipedia.org/wiki/Google_Bomb

ment, abstention or indifference, and disagreement respectively. For example, with the following HTML code the author of a page can link to a site in order to criticise it and, in doing so, communicate (to search engines but also to human readers) that she disagree with it.

⟨a rel=″vote-against″ href=″http://example.com″⟩bad site⟨/a⟩

Although, the proposal was made in early 2004, it seems the adoption of this microformat is minimal. By Nov 26, 2004, Technocrati had aggregated only 299 VoteLinks from 157 pages [3]. At this time, the sites of the two US presidential candidates received the greatest number of VoteLinks. An important point of VoteLinks proposal is that the specification is very simple so that humans can understand, remember and use it easily.

XFN (Xhtml Friends Network) [13] is another *Technorati.com* initiative which proposes another set of values for the rel attribute of the a HTML tag. Using these values authors can represent their relationships to other people on the Web such as, for example, friend, colleague, neighbor, sibling, spouse, date. This microformat is supported and adopted by the users of the popular blogging tool Wordpress (available at *wordpress.org*).

Another initiative has recently been proposed by *Google.com* and embraced by many other search engines and blogging tool creators. *Google* suggested to add rel=″nofollow″[4] to a hyperlink in order to indicate that the destination of that hyperlink should not be afforded any additional weight or ranking by user agents which perform link analysis upon web pages (for example, search engines). The motivation for this proposal is the problem of "blog spammers": automatic programs that post comments and links on blogs in order to manipulate PageRank into assigning a high rank to certain web site. *Google* proposes that blogging tools convert the links found in comments posted by all users by adding the rel=″nofollow″ attribute to the corresponding a tag. The rationale is that links appearing in comments to blog posts are not created by the owner of the blog and therefore should not improve the rank of the linked page by drawing upon the rank or importance of the blogger. The rel=″nofollow″ attribute instructs search engines not to consider the link as an expression of the opinion of the author of linking page. Thus, it is similar to rel=″vote-abstain″ in VoteLinks but has received greater recognition because of *Google*'s position as the most popular search engine on the Web. A drawback is that it does not allow the user to express a negative preference for a link.

All these proposals consider adding machine readable semantics to the existing linking structure of the web. In general, the Semantic Web project has been concerned with adding such semantics to the Web [2]. However, we note how many of the current Semantic Web initiatives seem to rely on a complex "top-down" approach. Semantic Web researchers are involved in defining precise and complex protocols to allow web authors to unambiguously express the content and the relations and semantics attached to the content. It remains to be seen whether the standards developed as part of this project will be adopted or whether web authors will continue to adopt simple incremental improvements. For example, the World Wide Web Consortium (W3C) recommended, on 27 June 2001, the XML Linking Language (XLink)[5]. The goal is to allow authors to describe links similar to the simple unidirectional hyperlinks of today's HTML, as well as more sophisticated links, such as asserting linking relationships among more than two resources, associate metadata with a link and express links that reside in a location separate from the linked resources. Its reduced adoption seems to suggest that the specification is too complicated to be understood and used by web authors. Instead, we think the initiatives we mentioned in this section go in the right direction, since they rely on simple incremental changes.

While the W3C is in charge of defining de jure standards for the Web, it is clear that certain companies (or even single individuals) have the clout to propose and consolidate new de facto standards. For example, *Google*'s rel=″nofollow″ attribute was widely adopted in few weeks. *Technorati* is pushing some very good initiatives but has been less successful in having them adopted.

Since most of the authors nowadays do not directly write HTML code, WYSIWYG editors should allow the author to express semantic values for the created link, for example they might give the user the possibility to "vote-for", "vote-against", "vote-abstain" with a simple click and possibly leave as default the "vote-abstain" or "vote-for" depending on the application.

Also browsers should show in some way the semantics of the link. There are many ways of doing this, such as with an extension or plugin for the browser or styling the site using CSS. To illustrate, we have modified the TargetAlert extension for Mozilla Firefox in order to show a thumb-up/thumb-down/thumb-null for rel=″vote-for/vote-against/vote-abstain″. The new extension, called SemanticLink, can be found at *http://moloko.itc.it/paoloblog/semantic_link.html*.

We should note that some researchers have previously attempted to use the text that appears close to the link (anchor text) to predict some information about the link itself, for example what is the topic of the linked page in the opinion of the author of the link [5]. *Google* uses link description to enrich the description of the linked page for indexing [3].

Of course having the "semantics" of the link unambiguously expressed by the author would allow more reliable predictions, results and new algorithms.

---

[3]A version of the page archived on Nov 26, 2004 is at http://web.archive.org/web/20041126091750/http://technorati.com/live/votes.html

[4]http://developers.technorati.com/wiki/RelNoFollow
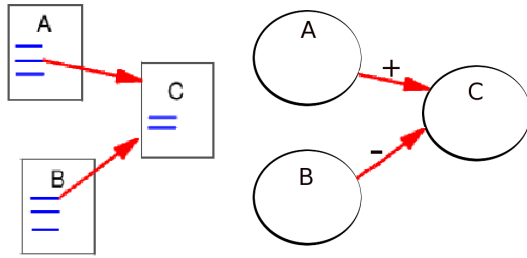
[5]http://www.w3.org/TR/xlink/

Figure 1. On the left, Web page A and Web page B link to Web page C. On the right, Epinions user A and Epinions user B judge Epinions user C. The graph structure is the same. Note that A trusts C while B distrusts C.

In short, we believe that enhanced Web intelligence requires that authors have the ability to describe their opinion on pages they link to. In this section we have presented the initiatives we judge as more promising in this direction.

4. Experiments

Although ranking algorithms may be able to find important pages (pages linked by many other pages and hence receiving a lot of attention), they cannot distinguish between positively and negatively endorsed pages. In this section, we present some experiments on real world data that demonstrate this intuition.

We describe firstly *Epinions.com*, the provider of the data, and then the experiments we have run on this data.

4.1 Epinions Data Set

Epinions.com is a web site where registered users can write reviews about products (such as books, movies, electronic appliances, restaurants, software, . . .) and also assign them numeric ratings. *Epinions.com* also allows the user to explicitly designate the users she trusts (i.e. reviewers whose reviews and ratings they have consistently found to be valuable) and the users she distrusts (authors whose reviews she finds consistently offensive, inaccurate, or in general not valuable). Based on this information, the system can provide personalized views to every user by boosting reviews of trusted users and suppressing reviews of distrusted users. For our experiment, we make a straightforward analogy with the web: web pages are *Epinions* users and links are trust and distrust statements (see Figure 1). The relevant difference is that *Epinions.com* allows to express positive (trust) and negative (distrust) links. Again the parallel with the VoteLinks proposal is straightforward: trust statements are vote-for links and distrust statements are vote-against links.

Let us now analyze the structure of the directed graph representing *Epinions.com* community. The data consists of 131828 nodes and 841372 edges. The edges are labeled either as 'trust' or 'distrust'. In Figure 1, trust and distrust statements are represented respectively by a + and a − near the edge. Precisely, 85.29% of the edges are labeled trust and only 14.71% are distrust edges. Some statistics about this graph can be found in [6]. In particular, the in-degree and out-degree distributions of this directed graph, treating both the trust and distrust edges, suggest a power law typical of the structure of the Web. However, the exponent of *Epinions.com* graph is -1.7 while, in general, the various power laws that have been observed on the Web present exponents below -2.0. For an analysis of the exponents of different networks see Table 2 of [9]. This graph presents other characteristics similar to the ones observed for the Web, for example the existence of a single large strongly connected component and the bow tie structure [4].

Moreover, the fraction of trust and distrust links confirm our intuition: humans tend to refer more to what they appreciate and less to what they don't appreciate. Of course, the user interface of *Epinions.com* has an heavy impact on this fraction as well. The other part of the *Epinions.com* dataset, not used in these experiments, provides additional evidence about this intuition. In fact, while users can freely rate items on a scale from 1 (minimum) to 5 (maximum), 77% of the ratings are 5 and 14% are 4. This human tendency to explicitly express opinions more on what they appreciate (in the *Epinions.com* example, both on other users and on items) implies that most of the time the assumption "a mention is an expression of appreciation" holds true. So, the assumption "a link from Web page A to Web page B is an endorsement of A on B", adopted by most of the search engines, is often correct. However, as we have already argued, this is not always the case and typical examples are political candidates sites or sites on highly controversial topics.

4.2 Experimental Setup

In this paper we claim that search engines could produce different rankings if they were able to consider the expressed intentions of web authors. At present, the link structure loses this information. In order to demonstrate our claim, we have run the PageRank algorithm against two versions of the *Epinions.com* graph. In the first setting, we have considered all the statements (both trust and distrust) as links between users. This is the information that can be expressed on the Web at the moment as there is no way in HTML to attach semantics to the link. In the second setting, we build a graph by keeping only the trust edges. In this way, we consider only the expressions of positive endorsement between users. In the following, the first setting is called *pr+/-* and the second is called *pr+*. In terms of the VoteLinks proposal, this would be the same as propagating reputation along both "vote-for" and "vote-against"

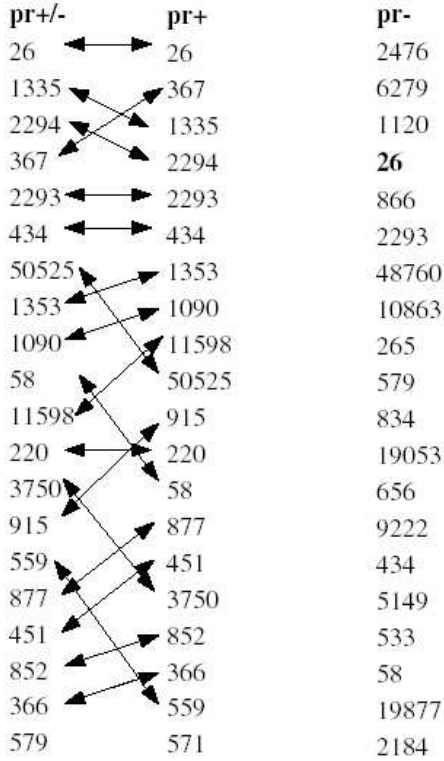| pr+/-	pr+	pr-
26	26	2476
1335	367	6279
2294	1335	1120
367	2294	**26**
2293	2293	866
434	434	2293
50525	1353	48760
1353	1090	10863
1090	11598	265
58	50525	579
11598	915	834
220	220	19053
3750	58	656
915	877	9222
559	451	434
877	3750	5149
451	852	533
852	366	58
366	559	19877
579	571	2184

Figure 2. Top20 ranked lists of Epinions.com users representing (left) "attention" and (center) "trust". The attention list is obtained running PageRank considering as links both trust and distrust edges, while the trust list is obtained running PageRank considering only trust edges. The rightmost list is obtained running PageRank considering as links only distrust edges.

($pr+/-$) or only along "vote-for" links ($pr+$). We have also run PageRank on the graph composed only of distrust edges ($pr-$). We will comment on this later in the section.

4.3 Results

The returned ordered lists are shown in Figure 2. The first two represent, respectively, *attention* ($pr+/-$) and *trust* ($pr+$). Note that, while it is possible to have a highly disliked page on top of the first list (as long as it received many links), this cannot happen on the second list that propagates only along the "vote-for" (positive) links.

User 367 is $4th$ in the attention list and $2nd$ in the trust list; this means that users 1335 and 2294 are more "spoken about" (considering both trust and distrust) while, if we consider only trust statements, user 367 is better placed. A search engine that wants to return the most appreciated users (or pages in the analogy) should probably return 367
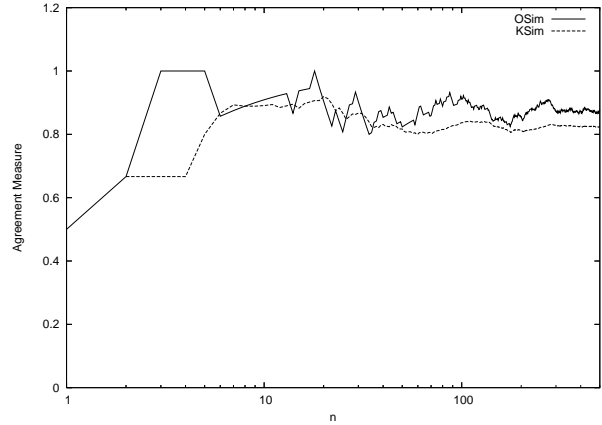

Figure 3. OSim and KSim of the two Top n lists generated by $pr+/-$ and $pr+$. The x-axis is in logarithmic scale to give more visibility to small values of n.

before 1335 and 2294. An opposite argument can be given for user 50525: she is $7th$ in the attention list but only $10th$ in the trust list. This means that, while she is known and judged (linked to) by many users, there is a non negligible portion of them that distrust her. Again, in this case, a search engine that wants to return mainly trusted pages should give preferences for example to 1353, 1090 and 11598.

We have also computed two measures of misalignment of the two lists as defined in [8]. The first one, denoted $OSim_n(\tau_1, \tau_2)$, represents the degree of overlapping between the top n elements of two rankings, τ_1 and τ_2. The overlap of the two sets A and B (each of size n) is defined to be $\frac{|A \cap B|}{n}$. The overlap measure $OSim$ does not give a complete picture of the similarity of two rankings, as it does not indicate the degree of agreement between the relative orderings of the top n users produced by two different rankings.

Therefore, we also use a variant of the Kendall's τ distance measure. For consistency with $OSim$, it is defined as a similarity (rather than a distance) measure, so that values closer to 1 indicate closer agreement. In short, $KSim(\tau_1, \tau_2)$ is the probability that τ_1' and τ_2' agree on the relative ordering of a randomly selected pair of distinct nodes $(u, v) \in U \times U$. The precise definition of $OSim$ and $KSim$ can be found in [8].

Figure 3 plots the two similarity measures when comparing the lists returned by $pr+/-$ and $pr+$, while Figure 4 consider $pr+/-$ and $pr-$.

4.3.1 Analysis of results

Figure 3 clearly shows the amount of disagreement between the attention list and the trust list. The information pre-
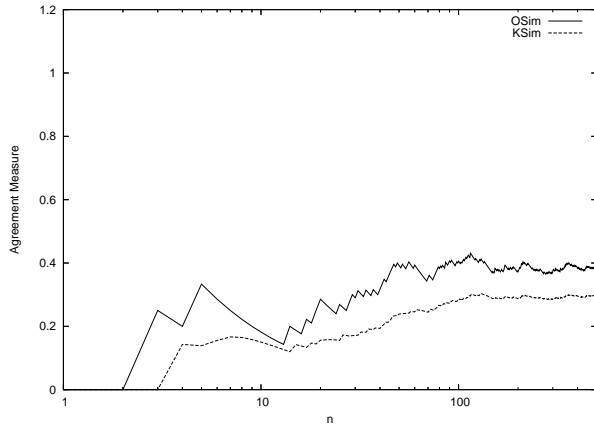
Figure 4. OSim and KSim of the two Top n lists generated by $pr+/-$ and $pr-$. The x-axis is in logarithmic scale to give more visibility to small values of n.

sented verbosely in Figure 2 is here presented in a compact form. The alignment tends to be a little bit greater than 0.8 (the difference between $OSim$ and $KSim$ is not large). We believe that this is an important fact. The difference between the attention list (what search engines can return at present) and the trust list (a list that would correspond to most appreciated pages) is significant.

When we are to compare the attention list with the list returned by $pr-$ (Figure 4), we see, as expected that the alignment is much less. While comparing Top n lists for small values of n, it is very small. However note that already for Top 50 lists, $OSim$ is around 0.4, indicating that the overlapping between the importance list and $pr-$ list is large enough. This means that in the attention list there are many users that are also in the list obtained by propagating reputation along negative links. This fact is far from ideal if a search would like to return only or mainly highly appreciated pages.

4.3.2   Distrust Links

Let us now analyze $pr-$ that is PageRank algorithm run on the graph composed only of negative links. In this case PageRank would propagate "negative" authority along links. But PageRank intuition is not really meaningful in this context. If user A distrusts user B and user B distrusts user C, this does not mean that user A would necessarily distrust user C. In some situations, the opposite can even be true. In short, propagating distrust does not make sense and the intuition behind PageRank totally loses its meaning in this context.

For this reason, $pr-$ Top 20 list (rightmost column of Figure 2) does not represent a list of most distrusted users and actually it is not clear what this list represents. However,

let us examine it. The users in the $pr-$ Top 20 list are very different from the users in the other two lists. The only user present in all of them is user 26, that happens to be an user with a very large number of incoming edges (both trust and distrust). In the real world, she would have been what we call a star, known by almost everyone, appreciated by many but also criticised by many. The other users in the $pr-$ list, instead, received mainly negative links. A short discussion about how the issues of utilizing "vote-against" information to produce more meaningful rankings is presented in Section 5.

5. Discussion

The ranked lists returned by $pr+$ and $pr+/-$ (Figure 2) are different and represent different concepts: trust and attention, respectively.

However, at present search engines can only return the attention list because trust and distrust statements cannot be unambiguously expressed in HTML. For example, during election time, search engines are able to return the candidates that are heavily discussed but cannot return the ones who are most trusted and appreciated.

We think that it is important to be able to have such a distinction and we support the *Technorati.com* VoteLinks proposal. Users should be able to express their opinion when they link to a page. Of course, the adoption of such a microformat would be easier and wider if HTML editors (and weblogging tools) were to make this feature available to normal users with a single click. We believe this is a correct first step to enriching links with semantics.

Since humans tend to mention what they like more than what they don't like, the $pr+$ and $pr+/-$ are largely overlapping and aligned. However, there are some users who have different positions in the two lists and this information is lost with the current HTML language.

Another point worth discussing is about $pr-$, that is applying PageRank to the graph composed just of distrust statements. The intuition behind PageRank in the context of negative links ("A negative authority is a page linked by many negative authorities") does not make much sense. The user distrusted by the most distrusted user is not necessarily the most distrusted. While trust is in some sense transitive, distrust is certainly not [6].

Social balance theory [11] states that "my enemy's enemy is my friend" but this is not really always the case. Our opinion is that if A distrusts B then B should not at all influence the opinions of A about other users or items, otherwise B could express her opinions in an instrumental way in order to influence recommendations to A: opinions of B should simply be ignored.

Instead, in this context, a simple adaptation of PageRank intuition is that a negative authority is a page judged as negative by many positive authorities. Hence, it makes sense

to propagate authority along vote-for links in order to discover positive authorities ($pr+$). Then using vote-against only in the last step, assuming that a page voted-against by many positive authorities is a negative authority.

However, this is still open to abuse. A site with a high positive PageRank is an authority - many pages link to it. Such a site has a lot of PageRank weight when it comes to assigning negative links. If this site negatively links to a smaller organisation, the smaller organisation will suffer in terms of its ranking. In the worst case scenario, a dominant organisation or interest group could use its positive PageRank to suppress opposition or competition. This type of behaviour could lead to negative link 'flame' wars where rival interest groups try to damage each others' rank.

However the goal of this paper is not to discuss how the semantic information encoded by authors along their links could be used but just to suggest to the Web Intelligence research community that a preliminary requirement must first be accomplished: extending the Web language so that semantic links can be expressed. [6] provides an analysis of different ways to propagate trust and distrust. Moreover, several algorithms have been proposed that exploit weighted relationships between nodes in a graph in order to infer authority and reputation, for example Trust Metrics [14, 1]. A logical next step for search engines is to produce ranked lists that are personalized to the opinions of the specific active user [7, 8]. In this context, links that express positive and negative preferences open new interesting possibilities.

Attention is not always a synonym of trust. We have shown that in fact, on a real world dataset, the two ranked lists representing these concepts are different. This fact clearly show that enriching HTML so that it is possible to represent some simple semantics along with links would benefit the research community and the web users at large.

6. Conclusions

In this paper we consider the potential for an extended linking language on the Web. Although search engines like *Google.com* tap the intelligence hand coded into the link structure of the Web, they are constrained from extracting more information from links by the lack of semantics available in the current linking model. We review several current proposals for extending the link mechanism and observe that any extension must be easy to use and be serviced by an intuitive tool set. Using a real world data set from *Epinions.com* as a a proxy for the Web, we demonstrate how additional link information would allow *Google*'s PageRank algorithm identify highly trusted web sites. We conclude that simple semantic extensions to the link mechanism would provide a richer semantic network from which to mine more precise Web Intelligence.

7. Acknowledgment

References

[1] P. Avesani, P. Massa, and R. Tiella. Moleskiing: a Trust-aware decentralized recommender system. In *Proceedings of the 1st Workshop on Friend of a Friend, Social Networking and the (Semantic) Web*. Galway, Ireland. 2004

[2] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, 2001.

[3] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *WWW7: Proceedings of the seventh international conference on World Wide Web 7*. Elsevier Science Publishers B. V., 1998.

[4] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the Web. *Computer Networks*, June 2000.

[5] S. Chakrabarti, B. Dom, P. Raghavan, S. Rajagopalan, D. Gibson, and J. Kleinberg. Automatic resource compilation by analyzing hyperlink structure and associated text. In *WWW7: Proceedings of the seventh international conference on World Wide Web 7*. Elsevier Science Publishers B. V., 1998.

[6] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, pages 403–412. ACM Press, 2004.

[7] T. Haveliwala, S. Kamvar, and G. Jeh. An analytical comparison of approaches to personalizing pagerank, 2003.

[8] T. H. Haveliwala. Topic-sensitive pagerank. In *WWW '02: Proceedings of the eleventh international conference on World Wide Web*, pages 517–526, New York, NY, USA, 2002. ACM Press.

[9] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

[10] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford, USA, 1998.

[11] H. Situngkir and D. Khanafiah. Social balance theory: Revisiting Heider's Balance Theory for many agents. Technical Report 0405004, Dept. Computational Sociology, Bandung Fe Institute, Indonesia, May 2004.

[12] Technorati.com. VoteLinks. http://developer.technorati.com/wiki/VoteLinks.

[13] Technorati.com. XFN (Xhtml Friends Network). http://gmpg.org/xfn/.

[14] C. Ziegler and G. Lausen. Spreading activation models for trust propagation. In *IEEE International Conference on e-Technology, e-Commerce, and e-Service (EEE'04)*, 2004.