LINK PREDICTION MODEL FOR PAGE RANKING OF BLOGS

S.Geetha
Research Scholar , Department Of Computer Science
PSGR Krishnammal College for Women , Coimbatore, India
Geethasellappanmca@gmail.com

K.Sathiyakumari
Assistant Professor, Department Of Computer Application
GRGovindarajulu School of Applied Computer Technology
Coimbatore, India
sathiyakumari@grgsact.com

Abstract - **Social Network Analysis is mapping and measuring of relationships and flows of information between people, organizations, computers, or other information or knowledge processing entities. Social media systems such as blogs, LinkedIn, you tube are allows users to share content media, etc. Blog is a social network notepad service with consider on user interactions. In this paper study the link prediction and page ranking using MozRank algorithm using blog websites. It finds out how all the websites on the internet link to each other with the largest Link Intelligence database. As link data is also a component of search engine ranking, understanding the link profile of Search Engine positioning. Here the MozRank algorithm is using backlinks from blog websites and linking websites quality. Good websites with many backlinks which linking the corresponding WebPage give highly value of MozRank. MozRank can be improved a web page's by getting lots of links from semi-popular pages or a few links from very popular pages. The algorithm for page ranking must work differently and MozRank is more comprehensive and accurate than Goggle's page rank. Another tool is Open Site Explorer that is ability to compare five URL's against each other. Open Site Explorer's Compare Link Metrics option is how one measures page level metrics, the other domain. This result can help to generate a chart form for the comparative URLs. A comparison chart of the important metrics for these pages is shown which makes it very clear and easy to compare the data between the five URL's.**

Keywords - weblog, media mining, Link Prediction, Page ranking, web mining and web usage mining.

I. INTRODUCTION

Online social network has disproportionate proscription and scheme for prospect out and encompass conceivable purpose. Like MySpace, Face book, LinkedIn, etc., MySpace is the most open, and user allowed search for and contact people beyond the intact network, whether they are devious members of the user social network or exhaustive strangers. Face book as a college social network application is much more proprietary and group oriented. Face book can only search for people that are in one of our substantiate networks. Those networks could embrace the company user work for, the college user attended, or even user high school. User can also join several of the thousands of smaller networks or groups that have been created by Face book users, some based on real-life organizations and some exist only in the minds of their founders.

1.1 Blog and Blogosphere

Blog is basically an online diary which usually comes in the frame of a website. Users can aggrandize postings from time to time. Blogs include photos, text etc. Blogs are in disproportionate templates, designs, background colors. Users can create own blog. There are websites like blogger.com, which is owned by Google. It allows users to choose blog address and control user's blog and post messages on to user's blog. Other web users can embrace to any user blog and user can edit or update or control the exposition. Blogs are in the form of personal diaries, new diaries, specific diaries some of them could be highly technical.

Blogosphere is a way of distinguishes the social creature that grows from a critical mass of blogs. Blogs are constructed, and with help the popular services like blog lines or technorati, blogs are ideally suited to interconnect with one another. Authors of a post may link within the post to another blog, they may comment on other blogs with links back to their own, or they may keep a blog roll of their favorite blogs. With such tight connections between blogs, information can be transmitted at an incredible pace. If one blog posts something that catches the attention of other reposting of that item, and the subsequent reposting of it by their readers means that the information can soon be all over the blogosphere.

1.2 RSS Feeds

RSS (really simple syndication) is a style through which users can easily read user web feeds. With RSS Reader can instinctive download stories, news, blog feeds to user desktop. For SEO, user can begin to really get user blog postings embrace by inducing facts and interesting blogs. This will result in giving your immersing, links and site immersion out to the masses. Typically most Internet users that receive RSS feeds utilize an RSS reader or news aggregator to keep up to date on their favorite websites or news resources, either from within their desktop, web-based device, or mobile-based device.

1.3 Track Back Ping

A track back also often called a link back, servers to let bloggers know when others have linked to their posts on their website. Track Back is a categorize of Peer to peer communication system that was construct to send declaration of streamlines between two web sites via a track back ping. “Ping (Packet Internet Grouper) is used primarily to troubleshoot internet connections. Ping in reference to Track Back citation to a small message sent from one web server to another. Track Backs are useful for squeaking a web server to another. Track Backs are useful for squeaking a web site that user have citation its web site within user own web site, and is prevailing with bloggers.

1.4 Social Network Analysis and Mining

Social Network Analysis and Mining (SNAM) is intended to be a multidisciplinary journal to serve both academia and industry as a main venue for a wide range of researchers and readers from computer science, social sciences, mathematical sciences, medical and biological sciences. Social network analysis and mining using different techniques from sociology, social sciences, mathematics, static’s and computer science.

The main areas covered by SNAM include: (1) data mining advances on the revelation and anatomizing of communities, personalization for solitary activities (like search) and social activities (like discovery of potential friends), the anatomizing of user behavior in open forums (like conventional sites, blogs and forums) and in commercial platforms (like e-auctions), and the associated security and privacy-preservation challenges; (2) social network modeling, construction of scalable, growth, and evolution patterns using machine learning approaches or multi-agent based simulation. Some analysis of social network analysis is: (1) content analysis (2) web metrics analysis (3) sentiment and affect analysis (4) video analysis. SNA techniques are used to examine web site and forum posting relationships. Various topological metrics (betweenss, degree, etc.) and properties (preferential attachment, growth etc.). There are several clustering (e.g., block modeling) and projection (e.g., multi-dimensional scaling, spring embedder) techniques to visualize their relationships.

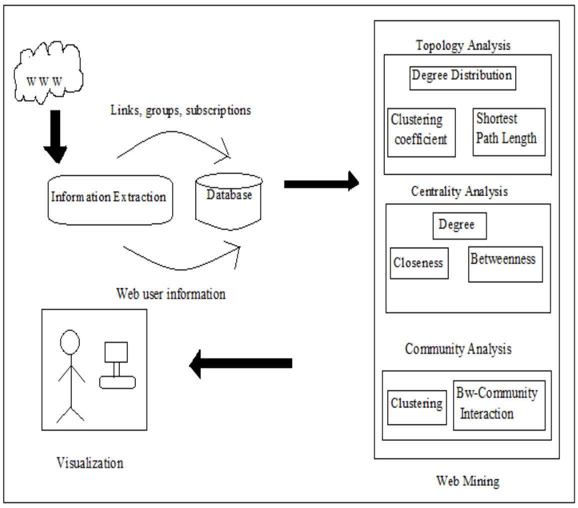

Figure 1: Analysis and visualization of Social Network Analysis (SNA)

1.4.1 Web Mining and Web Usage Mining

Web mining is the integration of information gathered by traditional data mining methodologies with information congregate over the World Wide Web. Web mining is used to extrapolate user behavior, appraise the persuasiveness of a discerning Web site, and benefaction quantity the treasure of a marketing campaign. Web mining allows us to look for patterns in data through content mining, structure mining, and usage mining. Content mining is used to examine data collected by search engines and web spiders. Structure mining is used to examine data related to the structure of a petulant web site and usage mining is used to examine data affiliated to a discerning user’s browser as well as data congregate by forms the user may have submitted during wed transactions.

Web usage mining discovers the user information access patterns from the web servers directly using automatic techniques. Different organizations collects and stores huge volumes of data from their web servers or from different sources and this data will be organized in a structured way to do their daily analysis to understand their users. Web usage mining is divided into two parts; the first one is transformation of data on WWW into appropriate domain, which includes preprocessing the data, identifying the transactions and integrating the data components. The second one is completely domain independent and involves application of general data mining techniques and pattern matching techniques. Cleaning of data is the primary step involved in the web usage mining process.

1.5 Link Predictions and Page Ranking

The link prediction problem is also pertinent to the problem of deduce missing links from a regard network. Number of domains in network, one domain constructs a network of companionship. This companionship based on perceivable data and then tries to extrapolate additional links. The basic task of link prediction in social and other networks is the explorations of improvised interactions that are transpire in the near future, among network members.

II. RELEATED WORK

Several studies have analysis of the link prediction and page ranking problems. The network analysis and the analysis the blogosphere in social medium, detecting blog user behavior also undertaken these work.

Predicting Response to Political Blog Posts with Topic Models: Tae Yano William W. Cohn Noah A. Smith, 2003. This paper describes the generation of the primary documents (posts) as well as the authorship and, optionally, the contents of the blog community's verbal reactions to each post (comments). Evaluate this model on a novel prediction task used to predict which blog users will leave comments on a given post.

Tracking Information Epidemics in Blog space: Eytan Adar and Lada A. Adamic, 2003 analyzed the pattern and dynamics of information spreading among blogs. The infection inference task is related to both link inference and link classification but makes use of non-traditional features unique to blog data. This paper task is to correctly label graph edges between blogs when one blog infects the other. Text similarity and URLs used to classify the one blog to another blog relation.

Social Networks and Reading Behavior in the Blogosphere: Tadanobu Furukawa, Yutaka Mastuo, Ikki Ohmukai, Kaki Uchiyama and Mitsuru Ishizuka, in 2007 describe an overview of social networks of weblogs integrated with analysis of users reading behavior and characterized the information diffusion on RR relations.

Link Prediction Approach to Recommendations in Large-Scale User-Generated Content Systems: In Nitin Chiluka, Nazareno Andrade, and John Pouwelse using popular Link Prediction algorithms for evaluate large dataset in online accurate alternative social networks. This link prediction algorithm is more scalable and accurate alternative to classical collaborative filtering in the context of UGCs. Implicit Structure and the Dynamics of Blog space:

Eytan Adar, Lada A. Adamic describe the information epidemics and create a tool to infer and visualize the paths specific infections take through the network. Here using new ranking algorithm, iRank, for blogs. This iRank algorithm acts on the implicit link structure to find those blogs that initiate these epidemics. Simple weighting scheme for the iRank algorithm to include the SVM to calculate the likelihood that one blog obtained information from another.

Content-based Modeling and Prediction of Information Dissemination: In this paper examine the link between information content and graph structure, proposing a new graph modeling approach, GC-Model, which combines both. GC-Model's top predictions covered covered 19% more of the actual future graph communication structure. It allows for the prediction of future thread structure. Link Prediction by De-anonymization: This paper describes the winning entry to the IJCNN 2011 social network challenge run by Kaggle.com. Here first demonstrate that partial crawls of a large real-world online social network can be effectively de-anonymized, whereas prior work studied de-anonymizing complete snapshots of social networks.

Ontology Engineering and Feature Construction for Predicting Friendship Links in the Live Journal Social Network: (Vikas Bahiwani, Doina Caragea) this paper address the problem of building interest ontology and predicting potential friendship relations between users in the social network Line Journal, using features constructed based on the interest ontology. It is organize user's interests in ontology and to use the semantics captured by this ontology to improve the performance of learning algorithms at predicting if two users can be friends. A hybrid clustering algorithm and combines hierarchical agglomerative and clustering paradigm builds the interest ontology.

The Spread of Media Content through the Blogosphere: In 2009 Juan Antonio Navarro Perez, Meeyoung Cha and Hamed Haddai described about the network structure and the spreading patterns of media content in the Blogosphere. Here find that user generated content, often in the form of videos or photos, is the most common type of content shared in blogs. Using spinner3dataset for used to extracting and predicting link structure.

Prediction and Recommending Links in Social Networks: In 2011 Lars Backstrom and Jure Leskovec developed a algorithm based on supervised random walks for existing missing interactions. This is combined the information from the network structure with node and edge level attributes. Using this algorithm to Face book social graph and large collaboration networks. Supervised Random Walks are not limited to link prediction and learning rank nodes in a graph. Link Prediction via Latent Factor Block Model: Ludovic Denoyer, Partrick Ganllinari. Sheng Gao (2011) addressed the problem of link prediction in networked data. The model can collectively capture globally predictive intrinsic properties of objects and discover the latent block structure, which shows the success of the success of the coupled benefits of latent feature based approaches and latent block models.

The Time Series Link Prediction Problem with Applications in Communication Surveillance: Zan Huang and Dennis K. J. Lin described the static graph representation in snapshot of the network is analyzed to predict hidden or feature links. This paper taking consider the temporal evolutions in link occurrences to predict link occurrence probabilities at a particular time. Combination of static graph link prediction algorithms and time series model produced significantly improved predictions than static graph link prediction methods.

Ethical aspects of web log data mining: The year of 2007 David L. Olson reviews the tools available, their uses in knowledge management and ethical aspects of blogs. This review has analyzed in three basic categories of application: first one is quantitative data; second one is behavior of web users, and finally massive quantities of text.

Tools to search the web were among the first applications of technology to web analysis. Mining Community Structure of Named Entities from Web Pages and Blogs: During that the year of 2006 Xin Li, Bing Lin and Philip S. Yu investigate the co-occurrences of named entities in such environment can be transformed into a graph clustering problem. This paper shows that the clustering algorithm can effectively discover interesting communities.

Predicting the Political Sentimental of Web Log Posts using Supervised Machine Learning Techniques Coupled with Feature Selection: (Kathleen T. Durant, Michael D. Smith) this paper illustrate the effectiveness of supervised learning for sentiment classification on web log posts. Using Naïve Bayes and SVMs on a novel collection of datasets created from political web log posts.

Present work: Other researchers have cruise weblogs for demarcate in writing phraseology pretending on link prediction in blogs. Yet this paper delineate the predict the link and ranking the pages in blogs using online tool, methods. This tool gives to very adequate results foe link prediction and page ranking in blogs.

III. METHODOLOGY

3.1 Online Tool

Data comes from the World Wide Web itself. Indexing large amounts of data is allowing relevancy research and finds continued activities in competitive community driven for general purpose to web scale search engine. Using the data (www.rte.in, www.independent.in, www.irishtimes.com, www.bbc.com/news, www.yahoo.com/news).

Using online tool for this paper implementation, the Majestic SEO (Search Engine Optimization) surveys and maps the internet and has created the largest commercial link intelligence database in the world. Link data is also a component of search engine ranking, understanding the link profile of user own and competitor is constantly revisiting web pages and around a billion URLs a day. Higher end features include deeper analysis and API access, allowing developers to integrate Majestic Data with their own toolsets.

3.2 Link Prediction Problem

The problem of link prediction has been generally in classification. Link prediction is a sub field of social network analysis. It is concerned with the problem of predicting the existence of links amongst nodes in a social network. The link prediction problem is interesting in that it investigates the relationship between objects, while traditional data mining tasks focus on objects themselves.

3.2.1 Link Popularity

Link popularity is a term that refers to how many other links point towards a particular website. The term link popularity also has two different forms, internal and external, which refer to the links coming from the websites own web pages and from other websites. Internal link popularity means the number of links to the website from web pages that belong to the particular website. External link popularity is the number of links from outside sources that lead back to the particular website. In the end, websites with high link popularity have what is called link cardinality or link superiority and have a reputation for being informative, as well as ranking highly on search engines.

Link popularity is also an approach that many search engines take when deciding where to rank websites. High popularity is an excellent way not only build a website but also show others how good the website it. When other web pages link to a particular website it draws additional traffic for the website, as well as giving it was what constitutes as votes in the search engine rankings. When two websites with very close levels of search engine optimization and information are being ranked by a search engine, more often the search engine will choose to rank the particular website with the higher link popularity first on search engine pages for engine results pages.

The inbound links from more obscure websites such as home pages for individuals and blog posts for individuals do not have as much impact on the search engine rank for the particular website or web page. The inbound links from major websites also carry more weight for the link popularity as a concept and not just a number, based on the idea that quality websites produce quality inbound links.

Search engines like Google use special link analysis systems to determine if link popularity for a particular website is worth ranking the website in a higher position. This is considered an off the page part of search engine optimization for advertising holders, as it is hard to influence the number of inbound links to a particular website without participating in link farming. While search engine optimization may be used to boost the popularity of the website, only quality information and attractive layouts can truly help with.

#	Page	External Backlinks	Referring Domains	Flow Metrics		Date
				Citation Flow	Trust Flow	
1	Title: RTÉ Ireland's National Television and Radio Bro... URL: http://www.rte.ie	54,687	4,683	62	66	7 Sep 2012
2	Title: http://www.rte.ie/player URL: http://www.rte.ie/player	17,689	1,880	63	59	6 Sep 2012
3	Title: RTÉ News, Breaking News from Ireland URL: http://www.rte.ie/news	42,727	1,280	59	57	7 Sep 2012
4	Title: Yacht once owned by Charles Haughey relaunched ... URL: http://www.rte.ie/news/2012/0812/haughey-yacht-...	1,205	808	28	23	20 Aug 2012
5	Title: Irish construction industry marks 62nd month of... URL: http://www.rte.ie/news/2012/0813/construction-i...	965	674	21	15	26 Aug 2012
6	Title: RTÉ Raidió na Gaeltachta URL: http://www.rte.ie/rnag	4,872	660	43	46	5 Sep 2012
7	Title: RTÉ Radio URL: http://www.rte.ie/radio	5,434	490	58	53	5 Sep 2012
8	Title: RTÉ Ireland's National Television and Radio Bro... URL: http://www.rte.ie/live	3,318	486	59	54	5 Sep 2012
9	Title: RTÉ News - News Headlines URL: http://www.rte.ie/rss/news.xml	4,087	470	40	44	7 Sep 2012
10	Title: RTÉ Radio 1 URL: http://www.rte.ie/radio1	3,500	448	41	36	5 Sep 2012

Figure 1: Top backlinks and Referring domains

3.2.2 Link Analysis

Link analysis is a data analysis technique used to evaluate relationships between nodes. Relationships may be identified among various types of nodes, including organizations, people and transactions. Network analysis, link analysis and social network analysis are all methods of knowledge discovery, each a corresponding subset of the prior method. Most knowledge discovery methods follow: Data preprocessing, Transformation, Analysis and Visualization. Link analysis is used for 33 primary purposes: (1) Find matches in data for known patterns of interest; (2) Find anomalies where known patterns are violated; (3) Discover new patterns of interest.

3.3 Ranking Authority

Domain authority represents SEOmoz's best prediction about how a website will perform in search engine rankings. Use domain authority when comparing one site to another or tracking the "strength" of user website overtime. Here calculate this metric by combining all our other link metrics (linking root domains, number of total links, MozRank, MozTrust, etc.) into a one single score. MozRank represents a link popularity score. MozRank is provided by seomoz and meaning value of page for engines.

It reflects the importance of my given web page on the internet. Pages earn MozRank by the number and quality of other pages that link to them. The higher the quality of the incoming links, the higher the MozRank.

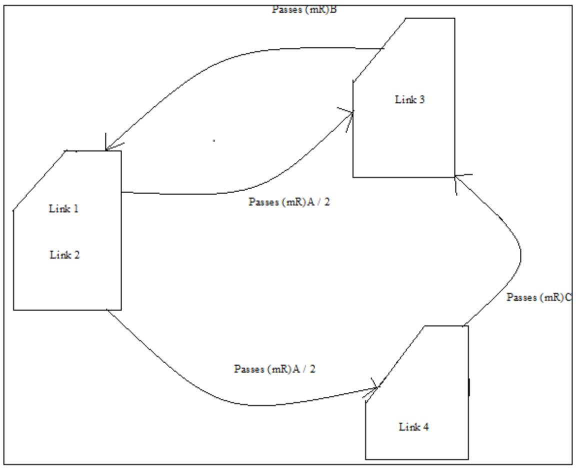

Figure 2: MozRank (MR) measure of global link authority

The performance is measured using three-time, leave-one-out, cross-validation. It is not worthy that here use no information of a direct relation between blog A and blog B when predicting the RR relation from A to B because to predict a possible RR relation when no recognizable relation exists between A and B.

3.4 Technical Definition of MozRank (MR)

MozRank refers SEOmoz's general, logarithmically scaled 10-point measure of global link authority (or popularity). MozRank is very similar in purpose to the measures of static importance (which means importance independent of a specific query) that are used by the search engines (e.g., Google's Page Rank). Search engines often rank pages with higher global link authority ahead of pages with lower authority. Because measures like MozRank are global and static, this ranking power applies to a broad range of search queries, rather than pages optimized specifically for a particular keyword.

The intuition behind MozRank is to leverage the democratic nature of the web. Every page has a vote and they can cast that vote by linking out to other web pages. Each time they link out all of the other links (votes) on the same page are slightly diluted. Thus, pages, which link to many other pages, aren't able to overwhelm the votes from pages that only link to a few other pages. Otherwise, if one page linked to the same page 1000 times, it would unfairly make that page rank artificially high in. The takeaway is that a given web page has only a quantifiable amount of link juice (ranking power) to spread via links (votes). Pages that receive a lot of links (votes) are considered more authoritative and are able to more authoritatively endorse pages they link to.

3.4.1 MozRank and MozRank Score

MozRank represents a link popularity score. It reflects the importance of any given web page on the Internet. Pages earn MozRank by the number and quality of other pages that link to them. The higher the quality of the incoming links, the higher the MozRank. MozRank calculate this score on a logarithmic scale **between 1 and 10**. Thus, it's much easier to improve from a MozRank of 3 to 4 than it is to improve from 8 to 9. An "average" MozRank of what most people think of a normal page on the Internet is around 3.

3.5 Metrics of MozRank

Three levels of metrics can be used. One is external MozRank, next one is domain level MozRank, and last one is MozRank passed.

3.5.1 External MozRank

Whereas MozRank measures the link juice (ranking power) of both internal and external links, external MozRank measures only the amount of MozRank flowing through external links (links located on a separate domain). Because external links can play an important role as independent endorsements, external MozRank is an important metric for predicting search engine rankings.

3.5.2 Domain-Level MozRank (DmR)

Domain-level MozRank (DmR) quantifies the popularity of a given domain compared to all other domains on the web. DmR is computed for both sub domains and root domains. This metric uses the same algorithm as MozRank but applies it to the "domain-level link graph". (A view of the web that only looks at domains as a whole and ignores individual pages) Viewing the web from this perspective offers additional insight about the general authority of a domain. Just as pages can endorse other pages, a link which crosses domain boundaries (e.g., from some page on searchengineland.com to a page on www.seomoz.org) can be seen as endorsement by one domain for another.

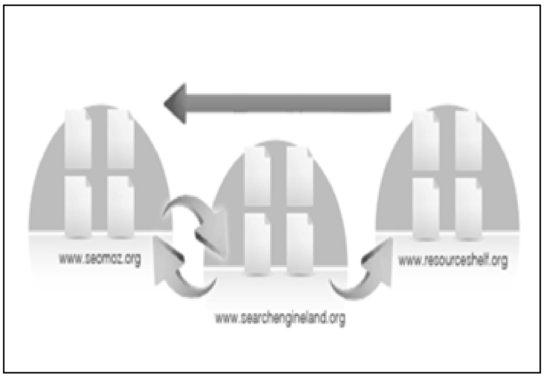

Figure 3: Domain-Level MozRank (DMR)

3.5.3 MozRank Passed

MozRank passed is the measurement of the amount of link juice (ranking power) a given link passes. Whereas, MozRank is the measurement of the link value of a individual web page, MozRank passed is the measurement of the value of an individual link. This is similar to dividing the amount of MozRank on a given page by the amount of total amount of links on the same page although MozRank passed takes into account some subtleties (duplicate links, link dampening, etc.) in order to increase accuracy.

3.6 Open Site Explorer

	Guest	Registered	PRO
Daily Reports	3	Unlimited	Unlimited
Number of Links	Up to 200	Up to 1000	Up to 10,000
Link Metrics	Top 5	Top 220	All

Table 1: Most Complete Link Data Available

In Open Site Explorer tool can be registered, it can able to access unlimited daily amount of reports, up to 10,000 links and all link metrics. Unlimited daily reports able to view up to 1000 links and the top 20 link metrics. A guest can access 3 daily reports, up to 200 links and the top 5 link metrics.

Quick overview of each measurement is displayed here:

(1) Page Authority – The “Page Authority” is scored out of 100. Based on an algorithmic combination of all link metrics, it predicts the page’s ranking potential.

(2) Domain Authority – The “Domain Authority” is also scored out of 100 and based on an algorithmic combination of all metrics however unlike the above, it predicts the domain’s ranking potential.

(3) Linking Authority – This “Linking Root Domains” is the number of unique root domains containing at least one link to the URL.

(4) Total Links – The “Total Links” is the number of all the links to the URL including internal, external, followed and no followed.

Feature of the Open Site Explorer tool is the ability to compare two URL’s against each other. A comparison chart of the important metrics for these pages is shown which makes it very clear and easy to compare the data between the two URL’s.

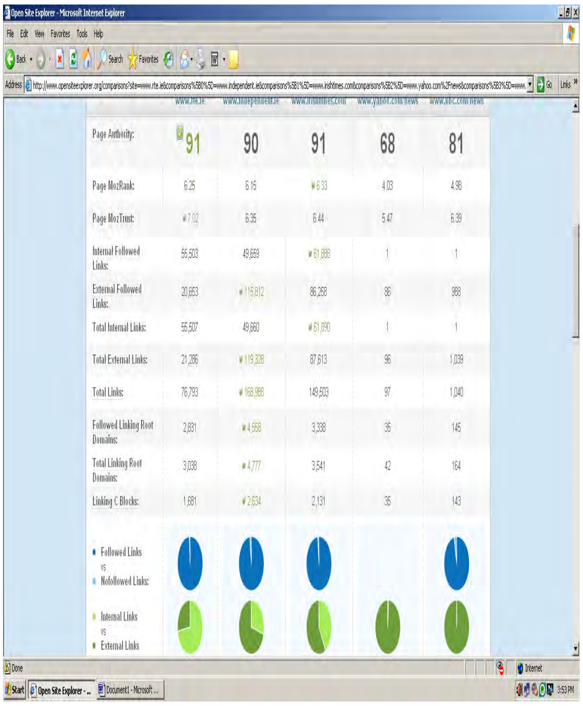

Figure 4: Comparison chart of important metrics of these pages

IV. CONCLUSION AND FEATURE WORK

A link prediction method that entrust only on local information to characterize propagatively the relationship between pairs of nodes is developed. In this paper, we introduce an link prediction model for page ranking in blogs using online tool. In this paper evaluate the bearing of comparing URLs with current updating for that URL links and also done evaluate the ratings that pages. SEO has been crawling the World Wide Web for mapping links between web sites. The index is now the largest salability available link intelligence database in the world. This internet map is used by SEOs, new media specialists, affiliate managers and online marketing experts for a variety of users surrounding online prominence including link building, reputation management, website traffic development, competitor analysis and news monitoring.

As link data is also a component of search engine ranking, benevolent the link profile of user, as well as competitor websites can endorse rational study of search engine positioning. The page rank algorithm remains edifice and quite badly known. Page Rank is only one of the criteria intricate in the Google search algorithm having good rankings! Here abomination is to pastime creating rich content for user, visitors, because it is the actual value added of user site.

As the scope of future enhancement of this research work, the following can be incorporated into the present work. Link data is a component of search engine ranking, understanding the link profile of user own, as well as competitor websites can empower rational study of Search Engine positioning. A comparison chart of the important metrics for these pages is shown which makes it very clear and easy to compare the data between the five URL's.

V. REFERENCES

[1] Admic and N. Glance, 2005. The political blogosphere and the 2004 U.S. election: Divided they blog. In Processing of the 2nd Annual Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics.

[2] M. Airoldi, D. Beli, S. Fienberg, and E. Xing. Mixed membership stochastic block models. JMLR, pages 1981-2014, 2008.

[3] Attardi G., Sinmi M. Blog Mining through Opinionated Words, In Proceedings of the 15th Text Retrieval Conference, 2006.

[4] Bo Pang and Lillian Lee. 2004. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. ACL Proceedings.

[5] K. Burton, A. Java, and I. Soboroff. The ICWSSM 2009 Spinn3r dataset. In Proceedings of the Third Annual Conference on Weblogs and Social Media (ICWSM 2009). San Jose, CA, May 2009.

[6] Chau, M. and Xu, J. (2007) 'Mining communities and their relationships in blog: a study of online hate groups', International Journel of Human-Computer Studies, Vol. 65, pp, 57-70.

[7] W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large scale social networks. In KDD'10, pages 10229-10338, Washington, DC, USA, 2010.

[8] A. Clauset, C. Moore, and M. E. J. Newman. Hierarchical structure and the prediction of missing links in networks. Nature, 453 (7191): 98-101, May 2008.

[9] W. Cukireski, B. Yang, and B. Hamner. How I did it: Will Cukierski on finishing second in the IJCNN Social Network Challenge. http://www.kaggle.com/blog/?p=728, 2010.

[10] David Liben-Nowell and Jon M. Kleinberg. The link prediction problem for social networks. In CIKM, pages 556-559, 2003.
[11] Domingos, P., and M. Richardson, Mining the Network Value of Customers, KDD’01: Knowledge and Data Discovery, San Francisco, CA, 2001.
[12] Dube, Jonathan: 2005. Blog Readership up 58% in 2004. CyberJournalist.net Jan 2005, www.cyberjournalist.net/news/001819.php.
[13] M. Gashler, C. Giraud-Carrier, and T. Martinez. Decision Tree Ensemble: Small Heterogeneous Is Better Than Large Homogeneous. In proceedings of ICMLA’ 08, pages 900-905. IEEE, December 2008.
[14] L. Getoor, N. Friedman, D. Koller, & B. Tasker. Learning Probablistic Models of Link Structure. Journal of Machine Learning Research, 2002.
[15] S. Herring, I. Kouper, J. Paolillo, L. Scheidt, M. Tyworth, P. Welsch, E. Wright, and N. Yu. Conversations in the blogosphere: An analysis “from the bottom up”. In Proc. HICSS-38, 2005.
[16] Kumar, R.; Novak, J.; Raghavan, P.; and Tomkins, A. 2003. On the Bursty Evolution of Blogosphere. In Proc of WWW.
[17] T. Lento, H. Welser, L. Gu, and M. Smith. The ties that Blog: Examining the relationship between social ties and continued participation in the wallop weblogging system. In 3rd Annual Workshop on the Weblogging ecosystem, 2006.
[18] D. Liben-Nowell and J. Kleinberg, “The link-prediction approach to anomalous email detection,” in SMC, 2006.
[19] [Marlow, C. 2004. Audience, structure and authority in the weblog community. In Proc. Communication Association Conference.
[20] Me, Liu C., Su H., Zhai C. A Probablistic Approach to Spatiotemporal Theme Pattern Mining on Weblogs. International World Wide Web Conference, 2006.
[21] Nardi, B.; Schiano, D.; and Gumbrecht, M. 2004. Blogging as a social activity, or would you let 900 million people read your diary? In Proc CHI 2004.
[22] Perseus Development 2004. The blogging iceberg: of 4.12 million hosted weblogs, most little seen and quickly abandoned. Technical report, Perseus Development.
[23] Shi, X.; Tseng, B.; and Adamic, L. A. 2007. Looking at the Blogosphere Topology through Lences. In ICWSSM.
[24] S. Wasserman and K. Faust. Social network analysis, Methods and Applications. Cambridge University Press, Cambridge, 1994.
[25] J. Yang and J. Lekskovec, “Patterens of temporal variation in online media,” in Proc. of the WSDM. ACM, 2011, PP. 177-186.