International Journal of Information & Computation Technology.
ISSN 0974-2239 Volume 7, Number 1 (2017), pp. 25-33

http://www. irphouse.com

Performance Analysis of Search Algorithms in Social Media

Steffy V. R and Dr. Rohini Venkat*

Department of Computer Science, Christ University, Hosur Road, Bengaluru, Karnataka 560029, India.

Abstract

Identifying beneficial information on the Social Media Networks is a challenging task in real time. This paper aims at providing an overview on various search algorithms across the social media networks. The task of inspecting through, to find useful information is colossal and this is where the role of Search Algorithms comes into picture. It presents a comprehensive approach to compare various algorithms' functionality based on their obligation in providing the result set. It also stretches an overview on performance of most popular search algorithms such as Page rank in Google, Edge rank in Facebook, etc. along with those social media networks which works without a search algorithm for example Twitter, Instagram.

Keywords: *Page-Rank, Edge-Rank, HITS, Feature selection*

INTRODUCTION

Social Media is the new arena for individuals to follow their desire, stay in touch and collaborate with each other. Social media has played a foremost role in developing a virtual sphere where users post their views, share/broadcast, and comment on various snapshots of existent contents and numerous porticos of the real world which in turn leads to a volatile propagation of social media on the Internet. It allows creation, retrieval and altercation of user-generated content that is ubiquitously accessible across

the globe. The heterogeneous and colossal data available in social media consisting of multimedia content, social context, users, geo-locations and other metadata information, this has resulted in numerous new research challenges and prospects.

LITERATURE SURVEY

Dr. Pushpa R. Suri and Harmunish Taneja [10] provides an insight on search engine optimization across World Wide Web by analyzing various search algorithms. It is a challenging task due to the expanding count of web users and the efficiency is influenced by user's perspectives and requirements. The paper proposes an integrated ranking model based on fused rank methodology which is a blend of popular and effective algorithms across the social networks.

PageRank: In Links & Law it describes as "PageRank relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of an individual page's value. Google interprets a link from page A to page B as a vote, by page A, for page B. But, Google looks at more than the sheer volume of votes, or links a page receives; it also analyzes the page that casts the vote." [1].

EdgeRank: In the site GeeksforGeeks "Facebook employs a machine learning algorithm that considers certain parameters to find relations between you and the person who has written that post. Prior to employing a machine learning algorithm, the EdgeRank Algorithm was being used by Facebook to rank the updates to be displayed on your feeds page. This algorithm not only ranks the feed but also sorts it to select which feeds should be shown on your feed at the very beginning and which one at the very last." [7]

SEARCH ENGINE OPTIMIZATION

Most of the web traffic is determined by the commercial search engine like Google, Yahoo, Bing, etc. Social media also play a vital role in generating visits to website and navigation to web pages for most Internet users. Despite of the services, content, products and information provided in the site. Search engines are exceptional in providing targeted traffic with respect to what web page offers and what is people looking for. Search engines are the thoroughfares to regulate the targeted traffic. If search engines fails in finding the site or add the content to their databases it would lead to lose of implausible opportunities to drive traffic. Search engines are working to improve their application or technology to crawl the web in depth and return better results.

SEARCH WITH ALGORITHMS

PageRank:

Google was the first main search engine to use an erudite algorithm to computing the relevancy of web pages in search results. PageRank relies on the uniquely democratic

nature of the web by using its vast link structure as an indicator of an individual page's value. PageRank is a link analysis algorithm which assigns a numerical allowance to each page in a hyperlinked set.

Page rank results from a "ballot" amongst the numerous pages transversely the World Wide Web regarding how important the page is. A hyperlink to a page counts as a vote of support. The PageRank of a page is defined recursively and depends on the number and PageRank metric of all pages that link to it.

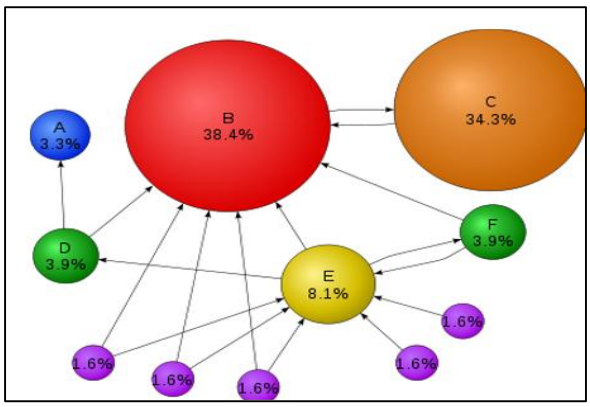

Fig 1: PageRank

Simplified Approach: Let us consider four web pages: Page A, Page B, Page C and Page D. If all these pages link to page A then the PR (PageRank) of page A would be the summation of the PR of pages B, C and D.

$$PR(\text{Page } A) = PR(\text{Page } B) + PR(\text{Page } C) + PR(\text{Page } D)$$

If page B as well has a link to page C and page D has links to all three pages. A page cannot vote twice and for this purpose it is considered that page B has given half a vote to each page. Similarly, only one third of page D's vote is counted for page A's PageRank.

$$PR(\text{Page } A) = PR(\text{Page } B)/2 + PR(\text{Page } C)/1 + PR(\text{Page } D)/3$$

In other terms, divide the PageRank by the total number of links that come from the page.

$$PR(\text{Page } A) = PR(\text{Page } B)/ \text{L}(\text{Page } B) + PR(\text{Page } C)/\text{L}(\text{Page } C) + PR(\text{Page } D)/\text{L}(\text{Page } D)$$

No page can have a PageRank of 0. Google performs a mathematical maneuver and gives every page a minimum of 1 - q.

$$PR(\text{Page } A) = [PR(\text{Page } B)/ \text{L}(\text{Page } B) + PR(\text{Page } C)/\text{L}(\text{Page } C) + PR(\text{Page } D)/\text{L}(\text{Page } D)\ldots.] \text{ q+1-q}$$

Complex Approach: A model of an arbitrary surfer who gets jaded after several checks and switches to a random page. The frequency of hits on the page by the random surfer determines the PageRank value of a page. A page which does not have any link to other pages If a page has no links to another pages, it becomes a sink and therefore makes this whole thing unusable, because the sink pages will trap the random visitors forever.

$$PR(p_i) = \frac{q}{N} + (1 - q) \sum_{p_{j \in L(p_i)}} PR(p_j)$$

EdgeRank:

EdgeRank is the algorithm of Facebook that decides which stories appear in every user's newsfeed. A story can be in a form of status update, comment, like, tag, event, and relationship status change or in a multimedia form such as photograph, video. This algorithm ranks the feed and also sorts it to select which feeds should be exposed on newsfeed initially and which are too be at the very last. The EdgeRank value is determined by the formula mentioned below:

$$\sum_{edges\ e} u_e\, w_e\, d_e$$

u_e ~ Affinity score between viewing user and edge creator

w_e ~ Weight of the edge type (create, connect, tag, like, comment etc.)

d_e ~ Time decay factor based on how long ago the edge was created

Affinity is a factor to measure the relationship between the creator of the story and the viewing user. The score is proportional to the closeness of the relationship. For instance, consider user A is interacting more with user B and less with user C then affinity score if high between user A and B. Affinity is one-way interacting significantly with another user which will prioritize stories and accordingly populate their stories in the users newsfeed. In this case stories of user B is populated in newsfeed of user A however not necessarily user A stories to appear on user B newsfeed.

Weight is the priority given to user post, based on the post type. The post type's hierarchy is followed as it garners more rendezvous in the network.

The hierarchy of post types on Facebook is as mentioned below.

Photos/ Videos > Links > Plain text status updates

Comments are weightier than likes but the overall weight of the post is affected by both type. A text-based update of status with m likes and n comments is more likely to occupy in the Newsfeed than a photo without any engagement.

Fig 2: Interaction with Author

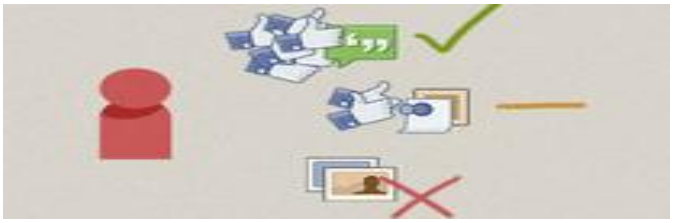

Fig 3: Interaction with the post type

Fig 4: Reactions from other users

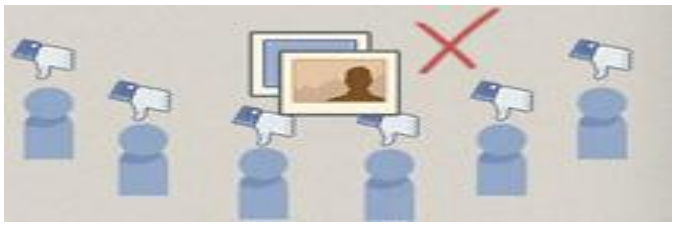

Fig 5: Negative feedbacks and Complaints

HITS Algorithm:

Hypertext Induced Topics Search (HITS) algorithm developed by Jon Kleinberg. HITS is applied on a sub-graph after a search is done on the complete graph. It uses a meek approach to document quality by assuming that if a document has a hyperlink to another document, then the former document accepts that the second document

contains valuable information. For instance, if site A contains references to many other sites, then the opinion of site A suggests that the sites mentioned are also good. It employs hubs and authorities to describe a recursive relationship between web pages. Algorithm applies two major aspects i.e. sampling component and weight propagation component. Sampling component constructs a focused collection of thousand web pages which is rich in authorities and weight-propagation component using the hub and authority weights determines the numerical estimates by an iterative technique.

Consider a page p has a hyperlink to page q, which is denoted as

$$p \rightarrow q$$

HITS algorithm computes hubs and authorities of the page A as follows:

$$a(p) = \sum_{p \rightarrow q} h(q)$$

$$h(p) = \sum_{q \rightarrow p} a(q)$$

NON-ALGORITHMIC SEARCH

Twitter has a complete stream of tweets in reverse chronological order. It is considered to be the perfect model by Facebook criticizers because all the content "gets seen" without the intrusion of an algorithm until Twitter Analytics became available showing that resultant reach amongst the noise of an unfiltered stream was shockingly low. The most adored of the social media platforms for engagement and reach is Instagram which is algorithm-free. Users appropriately hash tagging in Instagram should check follower numbers and reach rocket and undoubtedly their use on Twitter is well established similarly Instagram's search involves taking all elements such as locations, users, hashtags and media as interest. All of this is grouped together into sets which can be queried using extremely efficient set operations such as AND, OR and NOT. The results of these operations are efficiently ranked and trimmed to only the most relevant profile for a given query. It takes into account who you follow and who they follow in order to provide a more personalized set of results. This means that it is easier for you to find someone based on the people you follow.

METHODS AND MATERIALS:

Searching from wide range of entities makes it stimulating to associate to most relevant and right result. Search algorithms across the various social media networks aims at maximizing the number of visitors by assuring the result matching the user attributes such as profile, search pattern, feedback from other users etc.

The search algorithms across most popular social media networks are as follows

Table 1: Social Media Network and Search Algorithm

Logo	Social Media Platform	Search Algorithm
	Google	PageRank
	Facebook	EdgeRank
	YouTube	HITS
	Twitter	No Algorithm
	Instagram	

The comparison study executed for the analysis of performance in search algorithms across Google, Facebook, YouTube, Twitter and Instagram is tested by the testing tool named Charles Proxy Tool. "Charles is an HTTP proxy / HTTP monitor / Reverse Proxy that enables a developer to view all of the HTTP and SSL / HTTPS traffic between their machine and the Internet." [13]

Table 2. Time based analysis on search algorithms:

Social Media Platform	Response Time(Millisecond/s)
Facebook	0.6
Google	0.15
YouTube	10000
Twitter	0.80
Instagram	3000

RESULTS AND DISCUSSION:

Google gave the search results in the least response time and YouTube with maximum response time. The search algorithms with response time less than a second are plotted in figure 1.

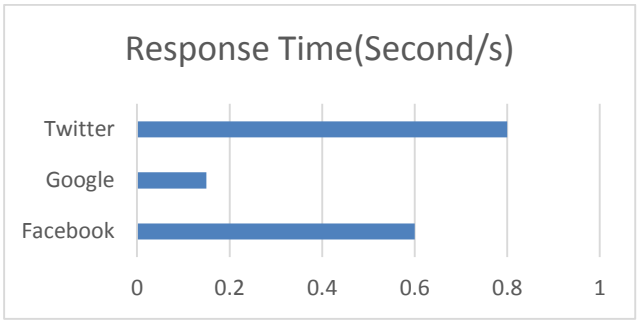

Fig 1: X Axis: Time in milliseconds; Y Axis: Social Networks

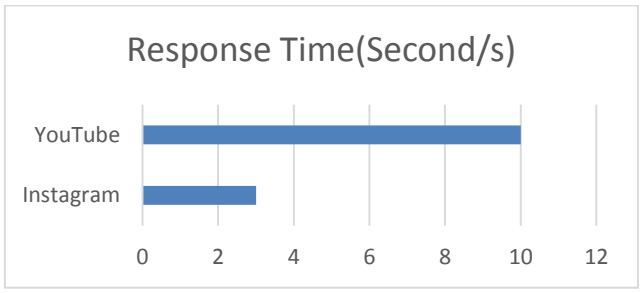

Fig 2: X Axis: Time in Seconds; Y Axis: Social Networks

CONCLUSION:

Social Media Monitoring extends prospects for more strategic approach. The scope of improvising the search algorithm is inevitable to meet the user expectation on the search results according to the platform if Facebook is more to find friends with more closeness or favorites, LinkedIn aims results based on Career perspective, similarly all other social media platforms are striving to enhance their search result optimization. In all the social media network, a prominent data flow for providing input is observed between the users and the search mechanism. To optimize the search the social networks should put in efforts to understand their users which in turn would increase the satisfaction of users and hence the maximum number of visitors or users to the application can be achieved.

REFERENCES:

[1] http://www.linksandlaw.com/technicalbackground-pagerank.htm
[2] https://moz.com/beginners-guide-to-seo
[3] http://searchengineland.com/guide/what-is-seo
[4] http://www.smartinsights.com/social-media-marketing/social-media-optimisation/social-network-algorithms/
[5] https://web.stanford.edu/class/cs54n/handouts/24-GooglePageRankAlgorithm.pdf

[6] http://computer.howstuffworks.com/google-algorithm.htm
[7] http://www.geeksforgeeks.org/edgerank-algorithm-facebook-news-feed
[8] http://sproutsocial.com/insights/facebook-news-feed-algorithm-guide
[9] http://mashable.com/2013/05/07/facebook-edgerank-infographic/#dazJJWWZqZqm
[10] https://thedigiterati.com/cracking-facebooks-news-feed-algorithm-new-definition-edgerank/
[11] http://www.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture4/lecture4.html
[12] https://webdocs.cs.ualberta.ca/~zaiane/courses/cmput499/slides/Lect14/sld055.htm
[13] https://www.charlesproxy.com/
[14] Dr. (Mrs.) Pushpa R. Suri and Harmunish Taneja, "An Integrated Ranking Algorithm For Efficient Information Computing in Social Networks", In International Journal on Web Service Computing (IJWSC), Vol.3, No.1, March 2012
[15] Jon Kleinberg, "Complex networks and decentralized search algorithms", In Proceedings of the International Congress of Mathematicians, Madrid, Spain, 20