Study of the Effectiveness of Social Media Search Algorithms

Anandhu Krishnan[1], Samson C Roy[2]
Student[1], Lecture[2]
Department of Computer Science Engineering
KMCT Engineering College
Corresponding Author's Email: -*anandhukrishna56@gmail.com*

Abstract

Finding relevant information on social media networks in real time is a difficult challenge. An overview of the different search algorithms used by social media networks is what this study aims to provide. Search engines are helpful since it takes a lot of labour to sift through data to discover relevant information. It offers a thorough way for contrasting the utility of different algorithms based on their responsibility to generate the result set. Additionally, it provides a summary of how the most well-known search algorithms, such as Google's page rank, Facebook's edge rank, and others, perform as well as social networking platforms without search algorithms, like Twitter and Instagram.

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

INTRODUCTION

People utilise social media to accomplish their objectives, stay in touch, and collaborate. Social media, which has contributed to the exponential rise of social media on the Internet, has enabled individuals to express their thoughts, share or broadcast them, and remark on multiple windows onto the actual world and diverse pictures of current events. It facilitates the creation, access, and editing of user-generated content that is made publicly available. The vast amount of data that is easily accessible via social media, including multimedia content, social context, users, geolocations, and other metadata information, has created a number of new study opportunities and issues.

LITERATURE SURVEY

In order to shed light on search engine optimization across the Internet, Dr. Pushpa R. Suri and Harmunish Taneja [10] examine a variety of search algorithms. It is a demanding task due to the growing number of internet users, and effectiveness is influenced by user views and needs. The study offers a fused rank approach-based integrated ranking model that combines well-liked and successful social network algorithms.

PageRank: According to Links & Law, "PageRank relies on the Web's inherently democratic character by using its enormous link structure as an indicator of an individual page's merit." A link from page A to page B is viewed by Google as a support for page B coming from page A. Google, on the other hand, examines the page that casts the vote in addition to the votes or connections a website receives. [1].

EdgeRank is used on the GeeksforGeeks website. Facebook uses a machine learning algorithm that takes into account a variety of factors to identify links between you and the post's author. Before adopting a machine learning system, Facebook used the EdgeRank approach to evaluate the changes that would appear on your feed page. This algorithm determines which feeds should be displayed first and last on your feed by scoring each feed and sorting the results. [7]

Search Engine Optimization

The majority of online traffic is determined by commercial search engines like Google, Yahoo, Bing, and others. Most Internet users heavily rely on social media to visit websites and increase website traffic. Regardless of the services, content, products, or information a website offers, search engines are exceptional at delivering focused traffic based on what users are looking for.

In order to route focused traffic, search engines serve as conduits. The loss of unexpected visitors will occur if search engines are unable to recognise the website or upload the information to their databases. In order to more thoroughly explore the web and deliver more accurate results, search engines are always working to enhance its application or technology.

SEARCH WITH ALGORITHMS

PageRank

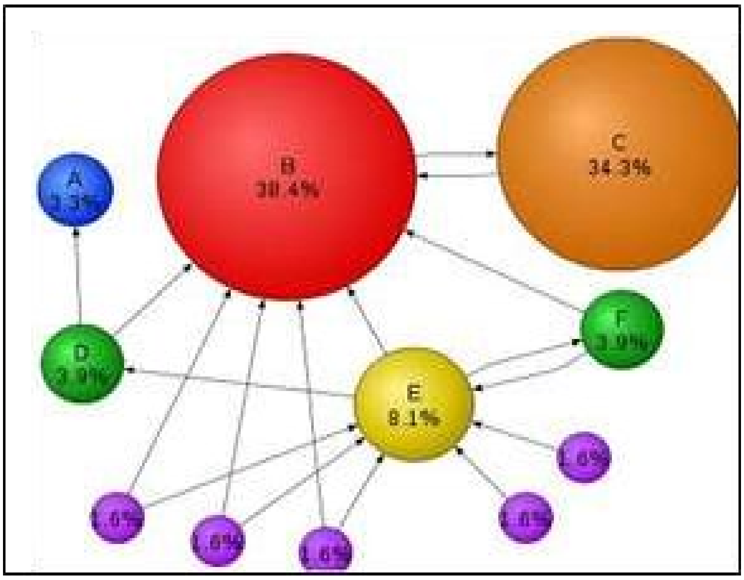

Fig 1: PageRank

Google was the first significant search engine to rank webpages in search results using a sophisticated algorithm. By employing the web's huge link structure as a signal of the significance of a single page, PageRank takes use of the web's particularly democratic character. PageRank is a link analysis method that gives each related page a numerical value. A "ballot" among the many pages that make up the World Wide Web chooses which ones are most important for a given page. An "approve" vote is represented by a link to a page. The quantity and PageRank metric of all the pages that connect to a page are used to recursively compute that page's PageRank.

Simpler Approach: Let's take a look at Pages A, B, C, and D of a website. The PR (page rank) of page A would equal the total of the PRs of pages B, C, and D if all of these pages link to page A.

PR(Page A) = PR(Page B) + PR(Page C) + PR(Page D)

Page D links to all three pages if page B and page C are likewise connected. A page may only vote once, hence page B is considered to have voted once for each page. Similar to this, page A's PageRank is determined using just one-third of page D's vote.

PR(Page A) = PR(Page B)/2 + PR(Page C)/1 + PR(Page D)/3

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

PR(Page A) = PR(Page B)/ L(Page B)+ PR(Page C)/L(Page C) + PR(Page D)/L(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(Page A) = [PR(Page B)/ L(Page B)+ PR(Page C)/L(Page C) + PR(Page D)/L(Page D)….] q+1-q

Complex Approach: A model of a random surfer who, after a few checks, grows jaded and navigates to a random page. PageRank is determined by the frequency with which a random surfer visits a page. a page that does not include any connections to other pages. If a website has no links to other sites, it becomes a sink, rendering the entire site ineffective because sink pages constantly trap random viewers.

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

EdgeRank

The Facebook algorithm known as EdgeRank selects which articles appear in each The stories that show up in each user's newsfeed are determined by the EdgeRank algorithm on Facebook. A status update, comment, like, tag, event, change in relationship status, or a photo or video can all be used to tell a story. The feeds are ranked and organised using this algorithm, which determines which feeds should appear first and last on the newsfeed. The EdgeRank value is calculated using the formula shown below:

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

ue ~ Affinity score between viewing user and edge creator

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

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

Affinity is a statistic that's used to assess the bond between the story's author and the reader. The score is related to how close the connection is. The affinity score between user A and user B is high, for instance, if user A interacts with user B more frequently than with user C. Affinity is a one-way interaction with another user that prioritises content and adds their stories to the user's newsfeed. In this case, user A's newsfeed contains stories from user B, but not necessarily stories from user A. Based on the type of post, the

priority given to a user's post is determined by its weight. The hierarchy of post types is maintained when additional network rendezvous occur. The hierarchy of Facebook post types is as follows:
Photos/Videos > Links > Status Updates in Plain Text Comments have more weight than likes, although both have an impact on the total weight of the article. A status update with m likes and n comments is more likely to appear in the Newsfeed than a photo with no engagement.

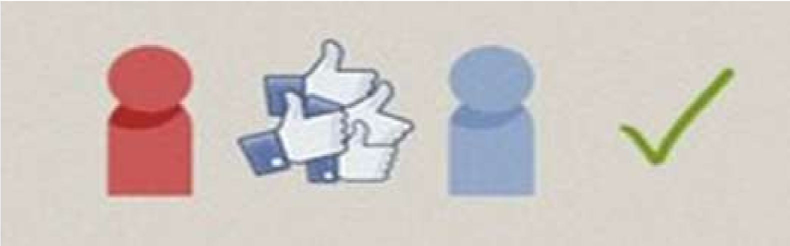

Fig 2: Interaction with Author

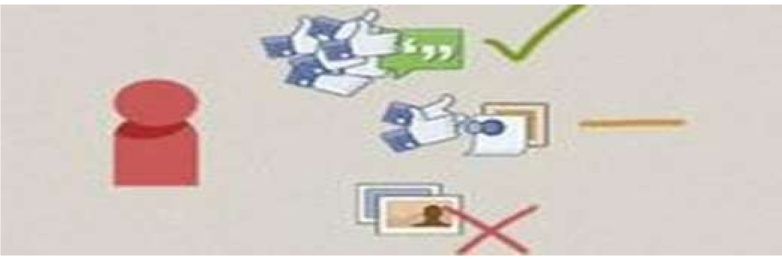

Fig 3: Interaction with the post type

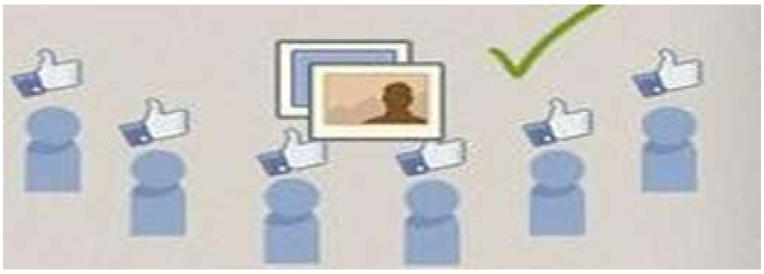

Fig 4: Reactions from other users

Fig 5: Negative feedbacks and Complaints

HITS Algorithm

Jon Kleinberg created the Hypertext Induced Topics Search (HITS) algorithm. After exploring the whole graph, HITS is applied to a sub-graph. If one document has a link to another, it is presumed that the first document recognises the validity of the second document. It includes information that is useful. For instance, if site A links out to many other websites, this suggests that site A thinks the websites being cited are likewise worthwhile.

It shows how websites employ hubs and authority to form a cyclic connection. The sampling and weight propagation are the two main parts of the approach. While the weight propagation component regularly generates numerical estimates using the hub and authority weights, the sampling component creates a targeted collection of thousands of authority-rich web sites.

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

There is a sizable reverse chronological stream of tweets on Twitter. Until Twitter Analytics became accessible, which showed that the resulting reach among the noise of an unfiltered stream was disappointingly low, it was seen by Facebook critics as the ideal approach as all information "gets seen" without the intervention of an algorithm. The most widely used social media platform in terms of reach and interaction is Instagram, which lacks algorithms. The follower and reach of those who properly utilise hashtags on Instagram will rise, and their usage on Twitter is indisputably widespread. In a similar vein, Instagram's search involves taking into consideration all factors, including locations, users, hashtags, and media.

The data is arranged into sets that may be searched using very efficient set operations like AND, OR, and NOT. These approaches yield results that are efficiently assessed and narrowed down to only the most relevant profile for a given query. It considers who you follow as well as who they follow to deliver a more tailored set of results. This means that depending on the individuals you follow, you will have an easier time locating someone.

METHODS AND MATERIALS

Finding the most accurate and relevant result when looking for a tonne of different topics is intriguing. By ensuring that the results meet the user's criteria, such as their profile, search activity, feedback from other users, and so on, search algorithms on social media networks attempt to maximise the number of visits. The most well-known social media networks employ the search engines listed below. Google, Facebook, YouTube, Twitter, and Instagram search algorithm performance is compared using the Charles Proxy Tool. In order to examine all HTTP and SSL/HTTPS communication between their workstation and the Internet, developers can use Charles, an HTTP proxy, HTTP monitor, or reverse proxy. [13]

Table 1: Social Media Network and Search Algorithm

<table>
<tr><th>Logo</th><th>SocialMediaPlatform</th><th>SearchAlgorithm</th></tr>
<tr><td></td><td>Google</td><td>PageRank</td></tr>
<tr><td></td><td>Facebook</td><td>EdgeRank</td></tr>
<tr><td></td><td>YouTube</td><td>HITS</td></tr>
<tr><td></td><td>Twitter</td><td rowspan="2">NoAlgorithm</td></tr>
<tr><td></td><td>Instagram</td></tr>
</table>

Table 2 Time based analysis on search algorithms:

Social MediaPlatform	ResponseTime(Millisecond/s)
Facebook	0.6
Google	0.15
YouTube	10000
Twitter	0.80
Instagram	3000

RESULTS AND DISCUSSION

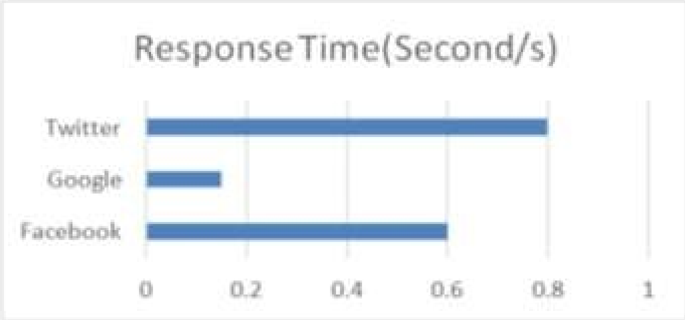

Fig 1: X Axis: Time in milliseconds; Y Axis: Social Networks

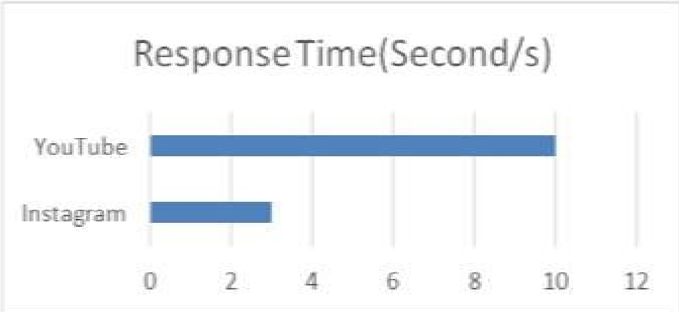

Fig 2: X Axis: Time in Seconds; Y Axis: Social Networks

The fastest search results came from Google, and the most time-consuming ones came from YouTube. Figure 1 displays the search algorithms with response times of under a second.

CONCLUSION

Opportunities for a more strategic approach can be found by monitoring social media. The need to optimise the search algorithm is unavoidable in order to satisfy user expectations for platform-specific search results. For instance, while LinkedIn aims to produce results from a professional perspective, all other social media platforms are similarly working to improve their search functionality. Across all social media networks, a sizable data flow for delivering feedback has been seen between users and the search function. Social networks should try to understand their users in order to improve search. This will boost user happiness and, as a consequence, the quantity of users or visits to the programme.

REFERENCES

1. http://www.linksandlaw.com/technicalbackground-pagerank.htm

2. http://www.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture4/lecture4.html
3. http://www.geeksforgeeks.org/edgerank-algorithm-facebook-news-feed
4. Dr. (Mrs.) Pushpa R. Suri and HarmunishTaneja, "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
5. http://www.smartinsights.com/social-media-marketing/social-media-optimization/social-network-algorithms/
6. http://mashable.com/2013/05/07/facebook-edgerank- info graphic/#dazJJWWZqZqm
7. http://searchengineland.com/guide/what-is-seo
8. https://www.charlesproxy.com/
9. https://moz.com/beginners-guide-to-seo
10. https://thedigiterati.com/cracking-facebooks-news-feed-algorithm-new- definition-edgerank/
11. http://computer.howstuffworks.com/google-algorithm.htm
12. https://web.stanford.edu/class/cs54n/handouts/24GooglePageRankAlgorithm.pdf
13. http://sproutsocial.com/insights/facebook-news-feed-algorithm-guide
14. https://webdocs.cs.ualberta.ca/~zaiane/courses/cmput499/slides/Lect14/sld055.h tm
15. Jon Kleinberg, "Complex networks and decentralized search algorithms", In Proceedings of the International Congress of Mathematicians, Madrid, Spain, 20

