ALGORITHM AND DATA STRUCTURES

ASSIGNMENT 3

FOUNDATIONAL ALGORITHM AND THEIR APPLICATIONS.

- What is a foundational Algorithm?
  Algorithms are basic step by step instructions, more like a math step for the computer to simplify tasks.
  For example, they are being used to simplify the complex of tasks we have been facing in the business and finance sector.
  Hill Top Algorithm:
- The Hilltop Algorithm was introduced in the year 2001 and was the first to use Sematics.
- Semantics usually mean deprecating the sounds, facial expressions, body language which the computer will generally fail to decode.
  How does Hilltop Algorithm function?

When you type in a keyword in google, the hilltop algorithm starts to find documents related to the keyword that we entered. The logic behind the algorithm is to work on special index of "Expert document". The expert documents are set of pages that link us to directories when we type into the World Wide Web. The algorithm was able to perform in a very small expert index of 2.5 million pages.

The algorithm consists of two broad phases.

1. **Expert Lookup :**
   An expert page is a page that deals with certain topics which are linked to many nonaffiliated topics.
   Two pages are called nonaffiliated when they have two different authors.
   Experts are identified as a sub set of pages that are scanned and crawled by a search engine.
2. **Target Ranking :**
   When a query is entered, we believe a page to be in authority if the best of the experts turn up in the query search.
   We have to carefully choose the text from a sub set of expert pages which are relevant to the query.

   This is how the algorithm delivers a high level of relevance in the data search process.
   Expert Pages:
   What makes an expert page different from a regular page?
   An expert page needs to be objective and diverse. It should contain recommendations that are unbiased and there has to be multiple nonaffiliated pages related to a search query. In

order to find experts, we have to find two recommendations or links pointing to the same organization.

Expert Documents:

1. Detecting Host affiliation:

We consider 2 hosts to be affiliated in case either one of following holds good:

- They share the same first 3 octets of the IP Addresses.
- The rightmost non-generic token in the host name is the same.

Tokens were considered to be sub strings of host delimited by. (Dot)

The suffix of the host name is to be considered generic. If this sequence occurs in a large number of distinct hosts.

Example: .com, .co, .uk.

Two hosts are considered to be affiliated if after removing the domain names, the right most token remains the same.

These tokens are named identifiers. To detect host affiliation, we use host affiliation look up. We use a union find algorithm and group hosts by assigning identifiers to non-generic tokens.

The host affiliation maps the set identifier to compare the host if and if they have the same value, they are considered to be affiliated.

2. Selecting the Experts:

This step includes processing the database of a search engine's pages, finding and selecting the set of pages which provide links to good sources of information on the query topic. This is done as follows:

We set a threshold for out-degree, consider 'k' (ex: k=5) we check to see if the search query or the URL points to more than k distinct non-affiliated hosts. We group all pages with out-degree greater than k, the threshold.

3. **Indexing the Experts:**

To locate expert pages that matches the user queries, inverted index is created to map the key words on the experts.

We store them at the following match points.

- Identify
- A code to refer to the kind of phrase
- The offset of the word in the phrase

Query processing:

1. Computing the expert score:

For an expert to be useful in a query, at least 1 URL containing all the keywords in the query must be there.

Score of an expert is a 3- tuple of form (s0, s1, s2)

Let k be the number of terms in the input query, q. The component Si of the score is computed by considering only key phrases that contain precisely k - i of the query terms.

Si = SUM {key phrases p with k - i query terms} LevelScore(p) * FullnessFactor(p, q)

Expert_Score = $2^{32} * S_0 + 2^{16} * S_1 + S_2$.

2. Computing the target score:

After computing the expert score, we consider the top n experts by ranking them and examining the pages they point to. These are called targets.

Advantages:

- The Hilltop algorithm focuses more on Data accuracy and not recovering the query.

EDGE RANK

There won't be 1.71 billion users using Facebook if Facebook wasn't as interesting as the algorithm it uses. Imagine getting the activity update of every move your friend does in her profile? Would you be interested in getting to know every activity she performed each day which will appear on your news feed?
Facebook came up with this amazing algorithm that helps filter and refine what you can see on your Facebook news feed and what will keep you scrolling through your wall.
The algorithm is named **Edge Rank** and here the word Edge stands for every action your friend takes on Facebook. The edge rank algorithm classifies each edge and then decides which edge will appear on your feed.

How it works?
Edge rank is calculated with the following attributes.

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

Where u_e is the affinity score.
w_e is the content weight and
d_e is the decay Rate
Though the complete algorithm remains a mystery, the steps of the algorithm were below.

1. Affinity Score :

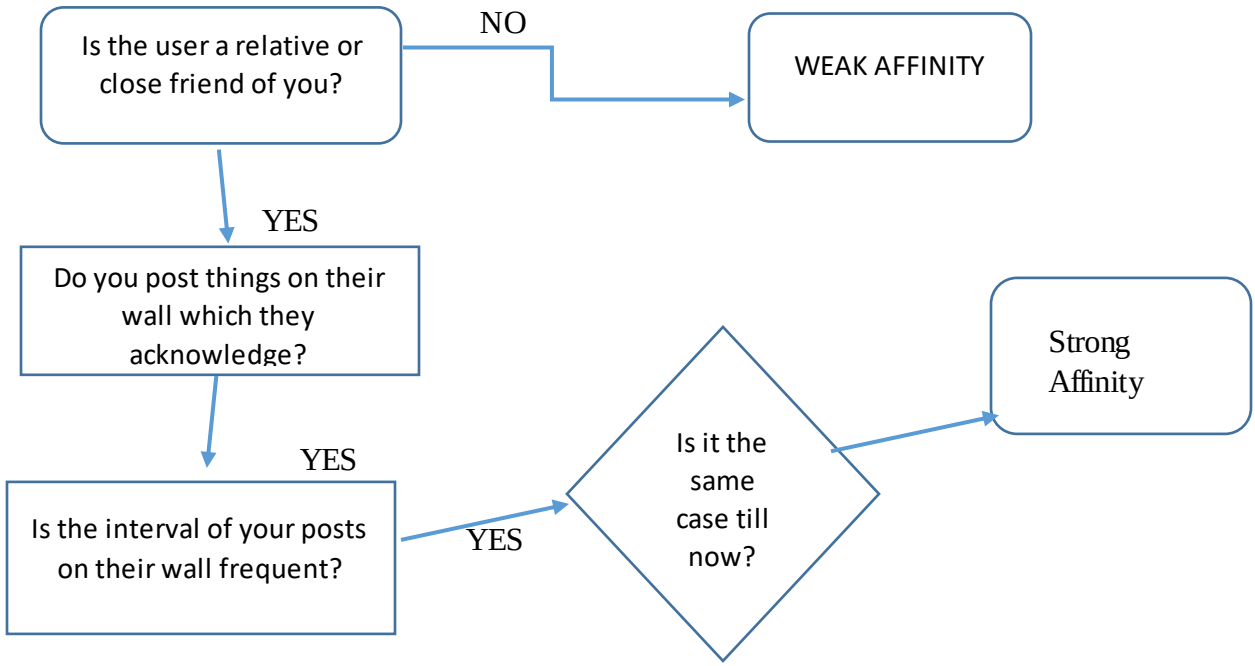
Affinity as the general word means, stands for the bond you share with a Facebook friend of yours.
There are again three factors which lead to concluding the bond of the user with that specific person.

- The strength of bond with the person
- How long ago was the action done
- The strength of the action that was made on that particular user.

For example, consider your affinity score with a friend who has been in your profile, on whose wall you actively post and you keep frequently posting on their wall.
Consider that the friend who frequently receives birthday post and photos with you on his/her wall, then the affinity of you and your friend will be determined as **STRONG**.

Commenting on the friend's post has a different weight when compared to just liking the same post. Hence, these factors sum up in determining the affinity.
Also, the frequency in which we tend to stay in touch with the person determines the affinity to a larger extent.
Say, the same friend to whose wall I have been actively posting, Zero posts in the present will reduce your affinity with the person.

A simple flow chart explaining affinity would be as follows.

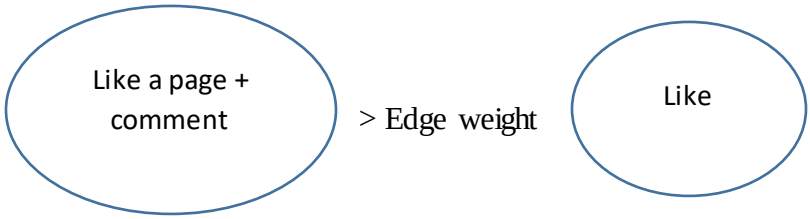


2. **Content Weight :**

The weights assigned to the edges depends on the activity being performed on the particular time. Say, comment edges have a higher weight when compared to like edges. If I comment on a page, it will hold much higher edge weight than just liking the page.

Added to this, the edge weights of the pages that I search for on Facebook and like has more edge weight than the ones I end up liking which appear in my wall.
Apparently, the newest of the Facebook features that has been just launched will have a very high edge weight when someone uses it.

The Edge weight can be pictorially represented as below.



3. **Decay Rate :**

The edge gets to become older and lose its value and come down the feed of the Facebook wall as time keeps progressing. To sum it up in mathematical terms, the decay rate is defined as **edge * 1/z** where edge is the action by a specific user and z is the time from the edge was created.

Advantages of Edge Rank:

- According to a recent survey, the post that you come up with is competing with 500 other posts to get into someone else's feed. This means that the traffic is highly monitored and the user gets to see customized version of how he wants it to be exactly.
- We see what we want to, eliminating things that we never want to come across.
- Without the algorithm, it is almost impossible for the average human to manually filter things he wants to see.