

# Graph Based Algorithms in Social Media Data Analytics

**Ms . Sajal Sharma**

*Student, Amity University Chhattisgarh,Raipur.*

**Mr.Advin Manhar**

*Assistant professor, Amity University Chhattisgarh,Raipur.*

## **Abstract:**

Increasing the number of users in social media sites and internet causing proliferation of data. Some structured ,unstructured, semi structured data generated by social media usually known as variety of data. Social media like Facebook, twitter, Instagram are producing a big amount of data through millions of collection.To forecast our advantage and needs online media information/data assumes a significant job.For dicover this analytical result the data of the social media has to be analysed efficiently.with the help of integrating graph theory algorithm with other analytical techniques we can analyse the the data of social media more efficiently.

## **Introduction**

As we know that accessive and regular use of social platforms ,internet and different websites is the major source for producing the data all over the world, this data termed as Big data. Structured , unstructured and semi structured are the formates of social media data ,also there are different ways to show social media data for analysis purpose. The structured data consist data in the form of table structure where as mixed media like audio, vedio and image data are are un structured data and the combination of both mixed media and table structure is known as semi

A series of graphs which captures dynamic properties of data called Big data. Big data also represents the properties and information of social media data . these data can also be present in the form of graphs as the representation of graph is

trouble free to solve tangled and complicated analytical problems. In this paper we discuss in data analysis how graph theory in computer science plays a very important role.

structured data.To compute the insightful outcomes fastly different calculations are applied to the online media information. For social media data analysis, graph theory and graph algorithm are the perfect way.graph theory concepts are one of most adopted technique ,its concepts and theorems are very helpful to solve many complications very easily. Further we discuss various graph analysis like optimal path analysis , path existing analysis, shortest path analysis and vertex centrality.

### 1. Graph theory

A graph consist vertices  $V$  and edges  $E$  , it is convinient to write a graph as  $G=(V,E)$  or a graph is a set of object connected with each other. Simple graph , directed and undirected graph, multi graph ,weighted graph , direct and undirected edges are some types of graph. We can show the graph in two different forms 1. Adjacency matrix 2. Adjacency list . adjacency matrix consist of rows and columns which are the vertices of the graph ,one of the row of adjacency matrix is with set of neighbour of the vertex on the other hand a collection of lists with its neighbour adjacency list , it contains only neighbour vertex as well as save time and space . some properties of graph are degree of the graph – it is its number of edges , (a) indegree – number edges entering inside the vertices and (b) out degree- number of edges emitting the vertices. Diameter property – is the length of the shortestpath .Rank property-is the quantity of edges , vertices and segments. Order property- It is its number of vertices.

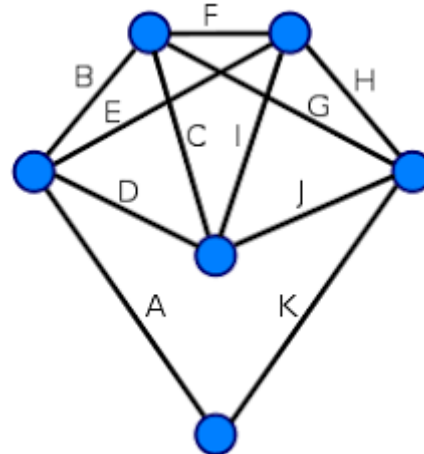
### 2. Application of graph in social media analytics

In 18th century graph theory is started with the swiss mathematician Leonard Euler who used the graphs to find the best path to cross over each of seven bridges for exactly once.

There are a few calculations which received chart hypothesis ideas in investigating web-based media information are determined underneath.

- Finding relative significance of site pages utilizing Page Rank calculation
- finding vicinity of vectors in diagram utilizing Random stroll with Restart calculation
- Finding associated hubs in chart utilizing associated parts calculation.
- Estimating width or span between vertices of diagram utilizing measurement assessment calculation.

Fig. Euler path graph



Later Hamilton develop a circuit that passes through each point exactly once but does not, in general, cover all the edges or Hamiltonian path is

a path in an undirected or directed graph that visits

each vertex exactly once.

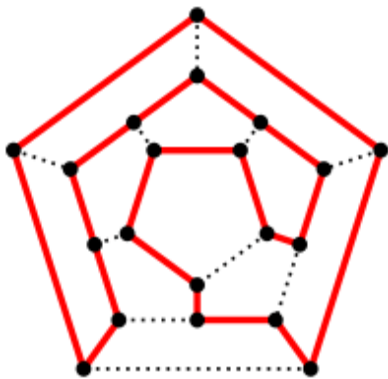


Fig. Hamilton graph

Yet, presently interpersonal organizations like face book and twitter are utilizing diagrams likewise charts are utilizing for discovering associations, networks, foundations and for positioning sites

### 3. GRAPH ALGORITHMS USED IN SOCIAL MEDIA ANALYTICS

It is extremely The information produced from online media immense which comprises of organized, semi organized and unstructured information.

Social media information can be spoken to as diagram for getting great scientific outcomes effectively and furthermore this chart portrayal is basic and proficient to use todissect web-based media information.

Various calculations are applied to decide different outcomes in investigation measure from the chart by anticipating diagramproperties.The most prominently utilized calculations to break down the web-based media chart are examined in the accompanying sub areas.

#### 3.1. Page Rank Algorithm on Twitter Data Set

Page Rank calculation is created to discover relative significance of pages dependent on the connections between the pages. Page rank of a page is chosen by finding the in connections and out connections of the page. In connection is the getting join from different pages of outer locales.

In the event that a page has more in connections, at that point the page get more votes and that page is treated as rank 1 page or most esteemed page. Out connection

is the connection indicated by the page to pages of outer destinations. This calculation is utilized in various applications like information mining, social network examination and web-based media information investigation. How page rank calculation is applied on twitter information for investigations is determined here.

In this calculation, the proposed work is influence(X) which predicts the quantity of individuals who will peruse a tweet and furthermore incorporates all retweets. On the off chance that an individual understands same message for multiple times then both are checked in light of retweets. X is an individual from adherents of Y then there is 1/following(X) likelihood for X peruses the tweet posted by Y. Following(X) is the arrangement of individuals that X Follows [7].

$$\text{Influence}(X) = \text{total} (1+p*\text{influence}(Y)/\text{Following}(Y))$$

where Y is the individual in Followers of X.

Utilizing this strategy impact of X is determined relying upon the impact of supporters of X.

### 3.2 Random Walk with Restart in Search Engines

Irregular stroll with restart (RWR) calculation is utilized to discover closeness between vertices of the graph[8]. Arbitrary Walk with Restart(RWR) is a calculation to gauge the nearness of hubs in chart. The nearness vector  $r_k$  from hub  $k$  fulfills the

equation[8]:

$$r_k = cMr_k + (1/c)ek$$

This calculation is utilized to rank the pages in web chart and to re-record the pages. Irregular stroll with restart calculation is for the most part utilized with Page rank calculation. RWR calculation is effectively executed and utilizing by Google. RWR calculation can be utilized in sight and sound diagrams to allocate watchwords to various sorts of media like pictures, recordings and sounds. RWR calculation is applied on

DBLP information to mine the examination networks to look through important meetings, creators and diaries.

### 3.3 Diameter Estimation Algorithm

Distance across is the one of the significant property of diagram which is utilized in various chart calculations. Breadth is the way length between two associated hubs of the diagram. Most extreme width is the exception. The successful measurement is characterized as least number of vertices needed to associate each combine of vertices of chart. This measurement assessment is

needed to comprehend and plan calculations for web charts and web-based media networks.

### 3.4 All-pair shortest paths Algorithm

The distance between two hubs in an interpersonal organization is a helpful component for some applications. For instance, it tends to be an element to anticipate most influencer in the organization and supporters of an individual. In this part an examination on the best way to inexact effectively the

distance between any two hubs in the chart in the public-private organizations is determined here.

This calculation is especially fascinating in finding ideal way and distance between two hubs. This way may change drastically regardless of whether we add a solitary edge. This calculation is exceptionally helpful to perform most brief way examination, ideal way investigation, way presence examination in dissecting online media networks and information.

## 4. CONCLUSION

Online media information and organizations contain enormous data as far as millions and billions. So the chart portrayal of this gigantic information disentangles investigation measure.

This paper examined about the diagram based calculations which are utilized in investigating social media information. At present page rank, RWR calculations are utilizing by google.

This paper momentarily clarified the use of page rank

also, RWR calculations for twitter information and for other network organizations.

In this way, for web-based media information investigation, chart based calculations are extremely proficient in discovering arrangements

in the investigation cycle. Further examination is needed to utilize chart calculations

## REFERENCES

[1] Facebook, "User statistics," February 2010. [Online].

<http://www.facebook.com/press/info.php?statistics>

[2] M. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.

[3] David F. Nettleton, "Data mining of social networks represented as graphs," *computersciencereview* 7(2013)1–34, Elsevier journal.

[4] T.h. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, *Introduction to Algorithms*, second ed., MIT Press and McGraw–Hill ISBN: 0-262-

53196-8, 2001.

[5] U Kang and Christos Faloutsos, *Big Graph Mining: Algorithms and Discoveries* Carnegie Mellon University fukang.

[6] U Kang, Charalampos E. Tsourakakis, and Christos Faloutsos, *PEGASUS: Mining Peta-Scale Graphs*, publication in *Knowledge and*

*Information Systems*

[7] <http://thenoisychannel.com/2009/01/13/a-twitter-analog-to-pagerank>

[8] Adams Wei Yu, Nikos Mamoulis, Hao Su, *Reverse Top-k Search using Random Walk with Restart* School of Computer Science, Carnegie

Mellon University, Department of Computer Science, The University of Hong Kong Computer Science Department, Stanford University.

[9] Osmar R. Zaiane, Jiyang Chen, Randy Goebel, *Mining Research Communities in Bibliographical Data*, *Advances in Web Mining and*

*Web Usage Analysis* pp 59-76.

[10] Suchit Pongnumkul, Kazuyuki Motohashi, *Random Walk-based Recommendation with Restart using Social Information and Bayesian*

*Transition Matrices*, *International Journal of Computer Applications* (0975 – 8887) Volume 114 – No. 9, March 2015.