IDENTIFICATION OF POTENTIAL INFLUENCERS IN FACEBOOK USING NETWORK GRAPH METRICS

[1]MURUGANANTHAM A, [2]G. MEERA GANDHI

[1,2]Research Scholar, Sathyabama University, Professor, Department of Computer Science and Engineering, Sathyabama University

---

Abstract- Social Media have been widely adopted by many businesses organizations. Social media sites like Facebook, Twitter are common phenomena for these companies to offer business services to their customers. Social networks are becoming a critical media for business and marketing. The amount of interaction generated at these sites are in the form like post, comments, Tweets, likes etc. influences the attitude and behavior of others. It is important to monitoring and identifying the potential influencer at these sites for better business opportunities and growth. For example, in the context of marketing on social networks, it is necessary to identify which users should be involved in an advertisement campaign. Each user is described by a large number of attributes, which transforms the problem of identifying relevant users in a needle in a haystack problem. It is important to quickly extract a limited number of meaningful characteristics that can be used to identify relevant users or potential influencers. Potential influencers may play a key role in the content dissemination and how users may be affected by different dissemination strategies. Network Graph Metrics such as Degree, Betweenness Centrality, Eigenvector Centrality, Clustering Coefficient, Closeness Centrality are used to identify the potential users in the social media networks. This paper presents a case study performed over a Facebook user with 250+ connections to explain how potential influencers are identified using Network Graph Metrics through visual representation. It is found those potential influencers are easy target for businesses to tap and keep them engaged for growth of their businesses.

Index Terms- Betweenness Centrality, Eigenvector Centrality, Facebook, Network Graph Metrics, potential influencer,

---

I. INTRODUCTION

Social Media site like Facebook is a huge virtual space where to express and share individual opinions, emotions, influencing one's view or opinions. Reviews and quality comments by influencer on the Internet are increasing their importance in the evaluation of products and services offered by business organizations. In some business sector like e-retailer, it is even becoming a fundamental variable in the “purchase” decision. A recent Forrester study showed that more than 30% of Internet users have evaluated products or services online. Consumers tend to trust the opinion of other consumers, especially those with prior experience of a product or service, rather than company marketing. An important customer's influence is called social influence influences other customers’ preferences by shaping their attitudes and behaviors. Monitoring, identifying and engaging the top influencers who are most relevant to the brand, product or campaign is become important now. We present a case-study to understand and identify potential users in a Facebook account using standard Network Graph Metrics and present our observations based on real social network data. The remaining parts of the paper are organized as follows. Section II describes previous literature of related research.

Section III describes various influence related Network Graph Metrics whereas Section IV presents a Facebook account case study and how to identify potential influencers and we conclude this paper in Section V.

II. LITERATURE OVERVIEW

Many techniques have been used to understand and monitor the users, who have much influence than others in the social media sites. Goyal et al. 2010 [1] propose models to compute social influence probabilities from real social network data. They use action logs of web data, and next define propagation of actions and propagation graph. Their algorithms first learn model parameters, and then test the learned models to make predictions. In their experiments, they calculate influence probabilities based on actions in action logs. Thus each action can appear completely either in training or test dataset.

King et al. 2009[2] summarize tasks and techniques in social computing mainly include but not limited to: social network theory, modeling, and analysis; ranking; query log processing; web spam detection;graph/link analysis and mining; collaborative filtering; sentiment analysis and opinion mining. Lots of prior researches focus on influence maximization problem [3][4], and polarity detection from web text[5][6], but few of them attempted to analyze deeply and find how a user sentimentally influences or is influenced by another in social networks. TwitterRank is proposed to identify influential users in Twitter [7]. As an extension of PageRank algorithm, it measures the influence taking both the topical similarity between users and the link structure into account. They truly process the tweets published in Twitter, and present their results to validate their solution on influence maximization problem.

A case study on a Facebook is presented to understand how potential influencers are identified using Network Graph Metrics using NodeXL.

III. INFLUENCE RELATED NETWORK GRAPH METRICS

A.A directed graph

Formally, assume a social network is modelled as directed graph G(V;E) where nodes V represent users, edges E represent social relationships among users and N represent size of network. Suppose user x adopts an innovation at time t1. We say that user x influences user y if and only if at time t2 when user y adopts the innovation, user x has already adopted it at an earlier time t1, at which time x and y were already friends. We therefore assume that social influence occurs when the information of a friend adopting the innovation has the potential to flow to neighboring nodes in the social network.

B. Social Influence

Social influence refers to the behavioral change of individuals affected by others in a network. Social influence is an intuitive and well-accepted phenomenon in social networks [8]. The strength of social influence depends on many factors such as the strength of relationships between people in the networks, the network distance between users, temporal effects, characteristics of networks and individuals in the network. ‘

Standard Network Graph Metrics such as centrality, closeness and betweenness are related to social influence in terms of the structural effects of different edges and nodes.

C.Edge Measures

Edge measures relate the social influence based concepts and measures on a pair of nodes.

Tie strength: According to Granovetter's seminal work [9], the tie strength between two nodes depends on the overlap of their neighborhoods. In particular, the more common neighbors that a pair of nodes A and B may have, the stronger the tie between them. If the overlap of neighborhoods between A and B is large, we consider A and B to have a strong tie. Otherwise, they are considered to have a weak tie. It is formally define the strength S(A,B) in terms of their Jaccard coefficient.

$$S(A,B) = \frac{|nA \cap nB|}{|nA \cup nB|} \qquad (1)$$

Equation (1) describes the strength of Tie, where nA and nB indicate the neighborhoods of A and B, respectively. Alternatively, it is called embeddedness. The embeddedness of an edge is high if two nodes incident on the edge have a high overlap of neighborhoods. When two individuals are connected by an embedded edge, it makes it easier for them to trust one another, because it is easier to to find out dishonest behavior through mu*tual friends*.

Edge Betweenness: It measures the total amount of flow across the edge. The information flow between A and B are evenly distributed on the shortest paths between A and B. edge betweenness leads to graph partitioning. By removing the edges of high betweenness scores to turn the network into a hierarchy of disconnected components or clusters of nodes. More detailed studies on clustering methods are presented in the work by Girvan and Newman.

D.Node Measures

Node-based centrality measure the importance of a node in the network. Centrality has attracted a lot of attention as a tool for studying social networks [12]. A node with high centrality score is usually considered more highly influential than other nodes in the network. Many centrality measures have been proposed based on the precise definition of influence. Centrality measures can classified into: radial and medial measures [12]. Radial measures assess random walks that start or end from a given node whereas the medial measures assess random walks that pass through a given node. The radial measures are further categorized into volume measures and length measures based on the type of random walks. Volume measures fix the length of walks and find the volume (or number) of the walks limited by the length. Length measures fix the volume of the target nodes and find the length of walks to reach the target volume.

Degree: It is radial and volume-based measure. The simplest and most popular measure is degree centrality.

Let A be the adjacency matrix of a network, and deg(i) be the degree of node i. The degree centrality C_i^{DEG} of node i is defined to be the degree of node:

$$C_i^{DEG} = \deg(i) \qquad (2)$$

Closeness: It is radial and length based measures. Unlike the volume based measures, the length based measures count the length of the walks. The most popular centrality measure in this group is the Freeman’s closeness centrality [13]. It measures the centrality by computing the average of the shortest distances to all other nodes. The closeness centrality C_i^{CLO} of node i is defined as follows:

$$C_i^{CLO} = e_i^T S1 \qquad (3)$$

Here S be the matrix whose (i,j)th element contains the length of the shortest path from node i to node j and 1 is the all one vector.

Node Betweenness or Betweenness Centrality: nodes of high betweenness occupy critical positions in the network structure, and are therefore able to play critical roles. It is often enabled by a large amount of flow, which is carried by nodes which occupy a position at the interface of tightly-knit groups. Such nodes are considered to have high betweenness. The betweenness centrality C_i^{BET} of node i is defined as follows:

$$C_i^{BET} = \sum_{j,k} b_{\frac{ijk}{b_{jk}}} \quad (4)$$

Here b_{ijk} is the number of shortest paths from node j to k, and bjk be the number of shortest paths from node j to k that pass through node i.

Eigenvector Centrality: It is defined as a function of number and strength of connections to its neighbors and as well as those neighbors' centralities. Let x(i) be the Eigenvector centrality of a node v_i. Then,

$$x(i) = \frac{1}{\lambda} \sum_{j=1}^{N} A_{i,j}\, x(j) \quad (5)$$

Here λ is a constant and A denotes the adjacency matrix. In nutshell, The Eigenvector centrality network metric takes into consideration not only how many connections a vertex has (i.e., its degree), but also the degree of the vertices that it is connected to.

IV. A FACEBOOK CASE STUDY TO IDENTIFY POTENTIAL INFLUENCER

A.A Facebook Dataset

It is well known that the number of connections in Facebook has risen dramatically in the last few years. First an experimental dataset from my Facebook account was downloaded. Considering our task need to identify potential influencers by applying Network Graph Metrics through NodeXL tool. We use 260+ users connection dataset in our experiments.

There are other network analysis tools like Pajek, UCINet, and NetDraw that provide graphical interfaces, rich libraries of metrics, and do not require coding or command line execution of features. However, we find that these tools are designed for expert practitioners, have complex data handling, and inflexible graphing and visualization features that inhibit wider adoption .

B. Network Graph Metrics

Facebook dataset, which is present in GraphML format is imported in NodeXL . The NodeXL represents the imported data set in the form of edge lists, i.e., pairs of vertices which are also referred to as nodes. Each vertex is a representation of an entity in the network. Each edge, or

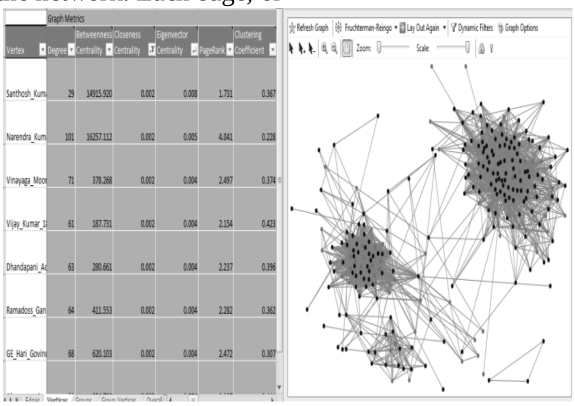

Fig. 1: Imported Facebook data in NodeXL

link, connecting two vertices is a representation of a relationship that exists between them. This relationship may be directed or not. Some relationships are bidirectional (like marriage); others can be uni-directional (like lending pen). Fig. 1 displays the imported Facebook data set.

NodeXL has a number of software routines for calculating Network Graph Metrics or statistics about individual vertices including degree, in-degree, out-degree, clustering coefficient, and closeness, betweenness, and eigenvector centrality. Fig. 2 generated Network Graph Metrics are summaries. The results of the Network Graph Metric calculations are added to the spreadsheet as additional columns that can be further combined and reused in Excel formula during analysis and visualization [15]. Spreadsheet features like data sorting, calculated formulae, and filters can be applied to network data sets directly. The network density is 8.5% of all possible ties.

Graph Metric	Value
Graph Type	Undirected
Vertices	261
Unique Edges	2903
Edges With Duplicates	0
Total Edges	2903
Self-Loops	0
Reciprocated Vertex Pair Ratio	Not Applicable
Reciprocated Edge Ratio	Not Applicable
Connected Components	9
Single-Vertex Connected Components	7
Maximum Vertices in a Connected Component	251
Maximum Edges in a Connected Component	2900
Maximum Geodesic Distance (Diameter)	8
Average Geodesic Distance	3.229963
Graph Density	0.085558503
Modularity	Not Applicable

Overall Metrics / Twitter Search Ntwrk Top Items

Fig. 2: Graph Metrics summary

C. Identification of Potential Influencers

We demonstrate how NodeXL tool identify potential influencers in a social media data set. NodeXL has Data Visualization feature By using this feature, the Facebook data is displayed as clusters through in-built clustering algorithms as show in the Fig. 3. Using the "Sub-graph Images" feature users can generate a set of thumbnail images that represent the connections around each node in

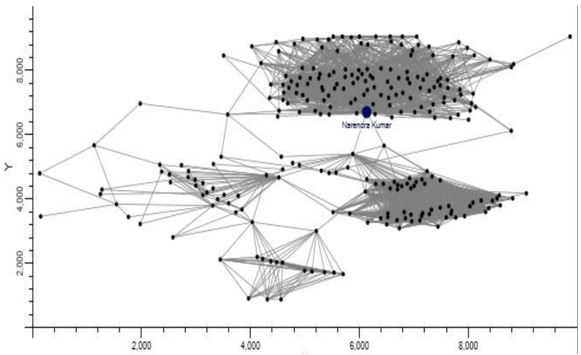

Fig. 3: Clusters or groups in Facebook dataset

The network. The Network Graph Metrics such as degree, Closeness Centrality, Betweenness Centrality and Eigenvector Centrality are calculated using (2), (3), (4) and (5). In Fig. 4, the nodes are sorted descending order of Betweenness centrality. The node, labeled Narendra is defined with highest Betweenness centrality. The vertex labeled Narendra having highest betweenness centrality (16257) can reach any information to the rest of users in the Facebook group very rapidly than the vertices having zero betweenness centrality. Among the vertices, the vertex labelled Narendra is most influencing person in the Facebook group.

In many cases, a connection to a popular individual (having highest Eigenvector Centrality is more important than a connection to a loner. The Eigenvector Centrality network metric takes into consideration not only how many connections a vertex has (i.e., its degree), but also the degree of the vertices that it is connected to. Fig. 5 shows the vertex labelled Santhosh is most popular in my Facebook network. The measure of Network Graph Metrics: Between Centrality and Eigenvector Centrality against Facebook users is shown in Fig. 6, which quickly identifies the potential influencers in my Facebook account of users.

Vertex	In-Degree	Out-Degree	Betweenness Centrality	Closeness Centrality	Eigenvector Centrality	PageRank
Narendra_			16257.112	0.002	0.005	4.041
Santhosh_			14915.920	0.002	0.008	1.731
Prabha_M			3783.135	0.001	0.001	2.717
Senthil_K			2984.555	0.001	0.000	1.818
Sundaresa			2620.258	0.001	0.015	1.252
Jayanthi_			2244.752	0.001	0.000	1.022
Selva_Kur			1237.374	0.001	0.000	2.240

Vertices / Groups / Group Vertices / Overall Metrics

Fig. 4: Highest Betweenness Centrality

		Graph Metrics					
Vertex	Tooltip	Degree	In-Degree	Out-Degree	Betweenness Centrality	Closeness Centrality	Eigenvector Centrality
Santhosh_		29			14915.920	0.002	0.008
Narendra_		101			16257.112	0.002	0.005
Vinayaga_		71			378.268	0.002	0.004
Vijay_Kun		61			187.731	0.002	0.004
Dhandapa		63			280.661	0.002	0.004
Ramadoss		64			411.553	0.002	0.004
GE_Hari_G		68			620.103	0.002	0.004

Fig. 5: Highest Eigenvector Centrality

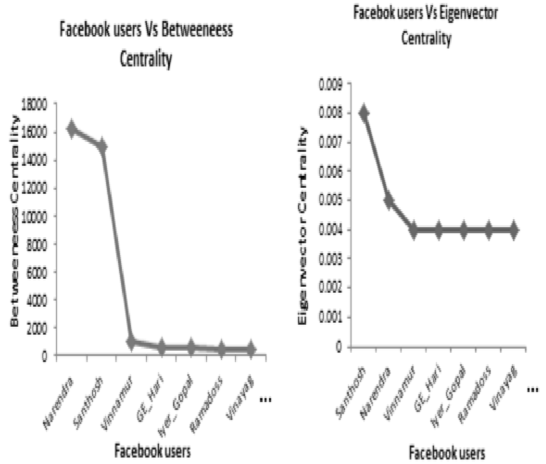

Fig. 6 Facebook users Vs Network Graph Metrics

CONCLUSION

In this paper, the potential influencers in the social media sites like Facebook are identified using Network Graph Metrics. NodeXL tool helps to generate and provide observations from a real social network (Facebook) dataset. Some of the important characteristics like Betweenness Centrality and Eigenvector Centrality are used to quickly classify and visually identify the potential influencers. We observed that highest value of measures: Betweenness Centrality as show in Fig. 4 and Eigenvector Centrality as shown Fig. 5 are real potential influencers in my Facebook account.

REFERENCES

[1] A. Goyal, F. Bonchi, and Laks V. S. Lakshmanan, "Learning influence probabilities in social networks", WSDM '10 Proceedings of the third ACM international conference on Web search and data mining, 2010.

[2] I. King, J. Li, and T. Chan, "A Brief Survey of Computational Approaches in Social Computing", IJCNN'09 Proceedings of the 2009 international joint conference on Neural Networks, IEEE Press, 2009.

[3] S. Bharathi , D. Kempe , and M. Salek, "Competitive influence maximization in social networks", Lecture Notes in Computer Science, 2007, pp. 306-311.

[4] J. Tang, J. Sun, C. Wang, and Zi Yang, "Social influence analysi in large-scale networks," Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 2009, pp. 807-816.

[5] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques", Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2002, pp. 79-86

[6] P. D. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews", Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL '02), pp. 417-424.

[7] J. Weng, E. P. Lim, J. Jiang, and Qi He, "TwitterRank: finding topicsensitive influential twitterers", WSDM '10 Proceedings of the third ACM international conference on Web search and data mining, 2010.

[8] D. Easley and J. Kleinberg. Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge University Press, 2010.

[9] M. Granovetter. The strength of weak ties. American Journal of Sociology, 78(6):1360–1380, 1973.

[10] M. Granovetter. Economic action and social structure: The problem of embeddedness. American Journal of Sociology, 91(3):481–510, 1985

[11] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America, 99(12):7821–7826, June 2002.

[12] S. P. Borgatti and M. G. Everett. A graph-theoretic perspective on centrality.Social Networks, 2006.

[13] L. C. Freeman. Centrality in social networks: Conceptual clarification. Social Networks, 1:215Ð239, 1979.

[14] Borgatti, S., Everett, M. G., and Freeman, L. C. (2006). UCINET 6, Analytic Technologies

[15] Shneiderman, B. and Aris, A. (2006). Network visualization with semantic substrates, IEEE Symposium on Information Visualization and IEEE Trans .Visualization and Computer Graphics 12 (5): 733-740.

★★★