

A Knowledge of Metrics on Analysis of Social Graphs

M. Vedhanayaki* and A. Kumaravel

AMET University, Chennai, Tamil Nadu – 603112, India; vedha.m7@gmail.com
Bharath University, Chennai - 600073, Tamil Nadu, India; drkumaravel@gmail.com

Abstract

Graph based data mining has become quite popular in the last few years. One of the best studied data structures in computer science and discrete mathematics are graphs. The use of social media has grown significantly in recent years. With the growth in its use, there has also been a substantial growth in the amount of information generated by users of social media. This paper discusses the metrics which are used in mining of social graphs.

Keywords: Data Mining, Graphs, Social Graphs

1. Introduction

Data mining is the analysis of a large amount of data to discover patterns or relationships or identification of correlations or patterns among dozens of fields in a database. Structured data mining is one of the data mining technique in which data are represented by structures. Graph mining is the special case of data mining. Graph mining has a strong relation with the multi-relational data mining. The structural nature of the data makes the intermediate representation and interpretability of the mining results much more challenging⁴. Social graphs are the network data illustrated using graphs. The analysis of social graphs is based on certain metrics.

The organization of the paper is as follows. Section 2 deals with the basics of graph theory which includes the representation of data through graph theoretic concepts. In Section 3 we discuss the various metrics of social graph mining. Section 4 gives conclusion.

2. Graph Terminologies and Representation of Data through Graphs

In this section we consider the base themes related to graph mining of social networks such as graph, social

networks representation using graphs. In general, a graph G is represented as $G(V, E)$ where V is a finite set of distinct elements and $E \subseteq V \times V$ is unordered pairs $\{u, v\}$ of distinct elements from V . Each element of V is called a vertex (point, node), and each element of E is called an edge (line, link). Edges of the form (u, u) , for some $u \in V$ are called self-loops, but in practical applications they typically are not contained in a graph.

2.1 Order of the Graph

The number n of vertices, (i.e. the cardinality of V), defines the order of graph and it is denoted by $|V| = n$.

2.2 Degree of a Vertex

The number m of edges (the cardinality of E), is called the size of the graph and denoted by $|E| = m$.

2.3 Derivatization Procedure

The number of neighbors of any vertex $v \in V$ in the graph identifies its degree.

2.4 Path

A sequence of connected vertices forms a path.

*Author for correspondence

2.5 Directed Graphs

The edges set is composed of ordered vertex (node) pairs; undirected if the edge set is composed of unordered vertex pairs. Moreover a graph G is

- Simple, if it has no loops or multiple edges.
- Acyclic if there is not any possibility to loop back again from every vertex; cyclic if the contrary holds.
- Connected, if there is a path in G between any given pair of vertices, otherwise it is disconnected.
- Regular, if all the vertices of G have the same degree.
- Complete, if every two distinct vertices are joined by exactly one edge.
- Bipartite, if the vertex-set can be split into two sets in such a way that each edge of the graph joins a vertex in first set to a vertex in second.
- A tree, if it is connected and it has no cycles. If G is a connected graph, the spanning tree in G will be a sub graph of G which includes every vertex of G and is also a tree.

How to represent a graph in computer memory is a key issue, due to the potentially high computational cost of many of the higher level operators we wish to perform. Two of the most popular data structures are ‘adjacency lists’ and ‘adjacency matrices’²⁵. Two other data structures, ‘incidence lists’ and ‘incidence matrices’ are similar to the former, with the distinction that the information stored indicates if edges and vertices are incident.

- Adjacency matrix - 1 indicates edge between 2 nodes
0 indicates no edge between 2 nodes
- Adjacency list - list of nodes connected to each node in some orders 5 min.

An example of a graph is given in Figure 1.

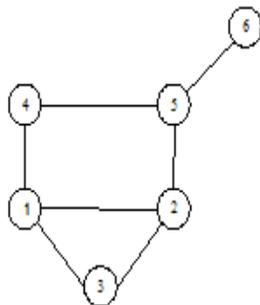


Figure 1. Graph G.

Table 1. Adjacency matrix of G.

	1	2	3	4	5	6
1	0	1	1	1	0	0
2	1	0	1	0	1	0
3	1	1	0	0	0	0
4	1	0	0	0	1	0
5	0	1	0	1	0	1
6	0	0	0	0	1	0

Table 2. Adjacency list of G.

Node	Connected to
1	2, 3, 4
2	1, 3, 5
3	1, 2
4	1, 5
5	2, 4, 6
6	5

3. Social Graph Mining Metrics

A social graph is a graph representation of social interactions and personal relationships in a network³. Social network analysis or social graph mining is the investigation of social structures, social position, role analysis, etc. Social graph mining involves the tasks such as centrality analysis, community detection, position analysis, network modeling, information diffusion, network classification and outlier detection, viral marketing and link prediction. Social networks are very large graphs which are defined by people who appear as nodes, and links which correspond to communications or relationships between these different people.

The analysis of social networks includes centrality analysis, community detection, position analysis, network modeling, Information diffusion, network classification, outlier detection, viral marketing, link prediction. The links in the social network can be used to determine relevant communities, members with particular expertise sets, and the flow of information in the social network. The mining of social graphs metrics are²:

1. Degree Centrality is the number of direct connections a node has.
2. Betweenness Centrality - is the Degree an individual lies between other individuals.
3. Closeness Centrality is the degree an individual is near all other individuals in a network (directly or indirectly).

4. Eigenvector Centrality is a measure of relative importance of a node.

5. Path Length is the distances between pairs of nodes in the network.

6. Structural equivalence is the level to which actors have a common set of linkages to other actors in the system.

7. Clustering coefficient gives a measure of the likelihood that two associates of a node are associates of themselves.

The problem of community detection in social networks is related to the problem of node clustering of very large graphs¹. In this case, we wish to determine dense clusters of nodes based on the underlying linkage structure. Social networks are a specially challenging case for the clustering problem because of the typically massive size of the underlying graph.

4. Conclusion

In this paper we overviewed the various measures involved in mining of social graph data. In an encour-

aging sense our highlighted measures provides some direction towards the social graph mining trends in order to reduce complexity.

5. References

1. Ferrara E. Community structure discovery in Face book. *International Journal of Social Network Mining*. 2012; 1(1):67–90.
2. Ferrara E, Fiumara G. Topological features of Online Social Networks. *Communications on Applied and Industrial Mathematics*. 2011; 2(2):1–20
3. De Meo P, Ferrara E, Fiumara G. A novel measure of edge centrality in social networks. *Knowledge-based systems*. 2012. DOI: 10.1016/j.knosys.2012.01.007.
4. Catanese S, Ferrara E, Fiumara G. Forensic analysis of phone call networks. *Social Network Analysis and Mining*.
5. Catanese S, De Meo P, Ferrara E, Fiumara G, Provetti A. Extraction and analysis of Facebook friendship relations. *Computational Social Networks: Mining and Visualization*. Springer Verlag.