NOVATEUR PUBLICATIONS
INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY [IJIERT]
ISSN: 2394-3696
VOLUME 5, ISSUE 5, May-2018

DATA MINING IN SOCIAL NETWORK ANALYSIS

MISS SUPRIYA KUMARI BISWAL (Author)
MCA Student P.E. S's Modern College of Engineering, Pune-05
supriyabiswal491@gmail.com

PROF. MRS. PRADNYA MULEY (Author)
Assistant Professor, P.E. S's Modern College of Engineering, Pune-05
pradnyamuley@gmail.com

ABSTRACT
In online communities and social software we can communicate with the web. Analysis of interactions happening is complex. Social network analysis (SNA) is a useful way to analyse. Data mining uses different type of statistical, machine learning and graphical methods and differentiates into a form of knowledge that is useful for most real-world applications. Analysis of social networks is now a popular field for research because it is useful for most applications. In this paper we have discussed the various data mining techniques used for social network analysis.

KEYWORDS: Social Networks, Data Mining, Analysis, Ego-centric analysis, socio-centric analysis, Social Network Analysis

1. INTRODUCTION

Data mining (DM) is a process of analysis / mine data from various sources and summarizes useful information. It is also called Knowledge Discovery Data and Knowledge Mining.Data Mining is a powerful tool that can help in finding patterns and relationships in our data. Data mining discovers information hidden from large databases Social Network Analysis is a research strategy that helps researchers find patterns of a set of actors. Artists are most people, but they can also be animals, organizations or nations. Problems with any research can be attributed to social networking analyses that have an measurable impact on the quality of the actor's relationship. Social network and its analysis is an important field and it has spread widely in the areas of data analysis. Social networks were researched through psychology, sociology, statistics and graph theory. The graphic theory based on the concept of social networks refers to the point of social relations of individuals and lines linking with them.(H 2010)

2. PAST REVIEW/STUDY

Recent research projects have been developed - methods of creating a statistical model of network data - computer science - related to the two closely related fields in machine learning and data mining. Examples of such data include data from social networks, network of web pages, complex connection databases, and interconnected people, events, locations, and text documents that are removed from documents. Such data sets are often referred to as "connectors" because it relates to intermediate relations between relationships (eg relationships between people, links in web pages or organized integrity between people and organizations).(D. Jensen 2002)

These algorithms are different from the more established set of data mining algorithms developed to analyse the installation data. For example, a proposal data set for learning medical diagnostic rules will represent each patient as a vector of the diagnostic test essay and the analysis will be assumed that if you know the disease of a patient, you will not tell about other patients. On the contrary, analysis of the relevant representation of similar data will retract the later assumption and add information about family relationships, workplace contacts, and other relationships in patients who influence their medical condition.(Wasserman S 1998)

Potential Possible Relative Model (PRM) (Friedman, Gater, Koller & Peeper 1999), Bessian Logic Program (BRP) (Kersting and De Ryder 2000), First-order, a small number of data mining techniques developed for temporary data, BleusianClassfears (Flat &Lachike 1999) and Relational Probability Tree (RPT) (Jensen and Naval 2002). The information of these cases can be learned directly from both statistical model design

and criteria data, facilitating the activity of data analysts, and improving the accuracy of the sulting model. The old method includes logic-based programming (ILP) (Magliten 1992, Dijersky and Liverpool 2001) and Social Network Analysis (Wassraman and Foster 1994).

3. SOCIAL NETWORK ANALYSIS AND DATA MINING

Data mining tools can be helpful in answering industrial questions that take a lot of time to solve a traditional way. Graphic mining methods can be used to minimize social network data mining using classifications / topologies, predictions, identities, performance, sample measurements, and metrics, modeling, data processing, evolution and structure, and communities.To extract the information represented in graphs one needs to

a) define metrics that describes the global structure of graphs,
b) find the community structure of the network, and
c) define metrics that describe the patterns of local interaction in the graphs,
d) develop efficient algorithms for mining data on networks and understand the model of generation of graphs.

Analysis of social networks can be applied to many issues and in-depth issues involving issues. Problems with any research may be influenced by the size of the relationship between the actor's relationship and hence social networking analysis can be beneficial.(Wasserman S 1998)

There are two types of social networking. "Sociocentric" or "Whole" network must have a relationship between all artists in one group. These artists can be members of a club, children's class, village or company's executive board. The center of the social network study is the group. There are people known by the "Egocentric" or "Personal" network, for example, a friend group or a specific team of a person in a company. (Xu G 2011)

SOCIO-CENTRIC (WHOLE) NETWORK ANALYSIS

- Emerged in sociology
- Involves quantification of interaction among a socially well-defined group of people
- Focus on identifying global structural patterns
- Most SNA research in organizations concentrates on socio-metric approach

EGOCENTRIC (PERSONAL) NETWORK ANALYSIS

- Emerged in anthropology and psychology
- Involves quantification of interactions between an individual.
- (called ego) and all other persons (called alters) related (directly or indirectly) to ego.
- Make generalizations of features found in personal networks.
- Difficult to collect data, so till now studies have been rare.

4. METRICS OF NETWORK

❖ CONNECTIONS

Homophily: The actor's limitations are similar to those that are incompatible with others. Equality can be defined by gender, race, age, occupation, academic success, status, value or any other special characteristics. Haemophilia is also known as evolution.(Carrington 2011)

Homeopathy is a principle that the contact between the same people occurs at a higher rate than different people. Extensive proof of homeopathy is cultural, behavioral, genetic, or physical information that flows through the network, it will be localized. Homeopathy has suggested that the interface between social features translates into the distance of the network, to connect with two people, one part of the information has to travel. This means that any social organization can rely on one of the important positions on the network for its transmission and it will be localized in social space and some fundamental dynamics will be followed because it communicates with other social elements of the social form environment.(Carrington 2011)

Multiplexity: The number of material-forms in the tie, for example, is that multiplexity 2 multiplexity connections with friends who are both people working together and working together will be related to the relational force.(D. Jensen 2002)

Multiplexity is the number of independent social connections between any two artists. This is defined as the "interface between relations and relationships" dialogue. Individuals have a single tie, such as a shared workplace, an unexpected relationship. The person has a tie multiplex during intervals, when those people interact in many social contexts. For example, there is a ball boss, and they have no relation outside their work, so their relation is unexpected, however, is a CSB co-worker and neighbor, so the relationships between B and C are multiplexes, because they interact with each other in various social roles.(D. Jensen 2002)

Mutuality/Reciprocity: The extent to which two actors reciprocate each other's friendship or other interaction.

Network Closure: Connector triads is a measure of perfection. The person's perception of closing the network (i.e., his friends are also friends) is called transit. Transitivity is the result of a person who needs cognitive closure or a situational thing. This "adherents of my friends" follow this principle.

❖ DISTRIBUTIONS

Bridge:A person, whose weak relationships relate to the structural hole, provides a unique link between two individuals or clusters. Due to the high risk of messaging or delivery failures, there is also a minimal path when there is now a short time impossible.(Carrington 2011)

Centrality: Centrality refers to a group of metrics that aim to measure the "importance" or "impact" (different senses) in the network of a particular node (or group). Examples of general methods of measuring "centralization" include NES concentration, proximity point, eigenvector centrality, alpha-cantee and degree centering.(D. Kempe 2003)

Density: The ratio of direct connections to the respective total number of related networks. If we compare it to two people and you realize that there are many artists in it which are not matched with other ("isolates") and most of the other populations are included in at least one color - you can probably conclude that the social population is too much in two populations It is different.(Domingos, Mining Social Networks for Viral Marketing 2005)

Distance: Minimal number of relationships required to connect two popular artists of Stanley Milgram's short global experiment and 'six degrees aside'.

Structural holes: There is no connection between the two parts of the network. Finding and Explaining Structural Hole One entrepreneur can offer competitive advantage. This concept is developed by sociologist Ronald Burt, and sometimes he is referred to as the alternative capital of social capital.(Domingos, Mining Social Networks for Viral Marketing 2005)

Tie Strength: Defined by the linear combination of time, emotional intensity, intimacy and reciprocity (depression). Strong bonding is related to homeopathy, equality and infection, whereas weak relationships are linked to the pool.

❖ SEGMENTATION

If every person is directly connected to each other with 'social circles', then it is known as 'clues', if the absolute accuracy is expected, there is a low strain of direct contact, which is inappropriate or structurally addable blocks.(J 2000)

Clustering coefficient: The node's two associate colleagues are likely to. Higher clustering coefficient shows more 'clickshit'.(J 2000)

5. CONCLUSION

The rise of social networks has a huge impact on the set of technologies developed for mining graph and social networks. Social networks are rooted in many sources of data and on many different aspects. Provides a efficient way to implement data mining and use the database. In this paper, we have briefly reviewed the

techniques used for social networking analysis and its applications for a little bit of time. SNA is a valuable contribution to HCI research because it allows for the study of complex patterns in online communication.

BIBLIOGRAPHY

1) Carrington, John Scott and Peter. "The sage Handbook of social network analysis." Sage Publications, 2011.
2) D. Jensen, J. Newille. "Paper of the Symposium on Dynamic Social Network Modeling and Analysis." National Academy of Sciences , 2002.
3) D. Kempe, J. Kelinberg, E. Tardos. "Maximizing the Spread of Influence through a Social Network." KDD, 2003.
4) D. Liben-Nowell, J. Kleinberg. "The link Prediction Problem for Social Networks." CIKM, 2003.
5) Domingos, P. "Mining Social Networks for Viral Marketing." IEEE, 2005.
6) Domingos, P. "Mining Social Networks for Viral Marketing." IEEE, 2005.
7) Freeman, Linton C. "A study in the sociology of science ." Empirical Press, 2004.
8) H, Du. "Data mining Techniques and Applications an Introducation." Cengage Learning Edition, 2010.
9) Han J, Kamber M. "Data mining: Concept and Techniques." Morgan Kauffmann, 2006.
10) J, Scott. "Social Network Analysis - A Handbook." SAGE Publications, 2000.
11) Kadushin/s, Charles. "Theories, Concepts and findings." Oxford University Press, 2011.
12) Kimberly A. Fredericks, Maryann A. durland. "The historical evolution and basic concept of Social Network Analysis."
13) L.Page, S. Brin. "The anatomy of a large scale hypertextual Web Search Engine."
14) M. Richardson, P. Domingos. "Mining Knowledge- Sharing Sites for Viral Marketing ." KDD, 2002.
15) MuxViz. "A tool for multilayer analysis and visualization of networks." Journal of Complex Networks.
16) P, Mika. "Social Netowrks and the semantic web." springer, 2007.
17) P. Domingos, M. Richardson. "Mining the Network Value of Customers." KDD, 2001.
18) PAge, S. Brin and L. "The anatomy of a large scale hypertextual web search engine."
19) Robert A. Hanneman, Mark Riddle. "Introduction to social network method: Riverside." 2005.
20) S. Chakrabarti, B. Dom, D.Gibson. "Mining the link structure of the world wide web." IEEE.
21) Scott, John. "Social Netowrk Analysis." 2000.
22) Stanley Wasswerman, Kathering Faust. "Methods and Applications." Cambriage University Press, 1994.
23) Valente, Tom. "Models, Methods and applications." Oxford University Press, 2010.
24) Wasserman S, Faust K. "Social Network analysis - Methods and Applications." Cambridge University Press, 1998.
25) Xu G, Zhang Y, Ki. L. "Web mining and Social Networking Techniques and applications." Springer, 2011.
26) Zafarani R, Abbasi MA, Liu H. "Social Media Mining an Introduction." Cambriage University, 2014.