

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology



ISSN 2320-088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 6, Issue. 6, June 2017, pg.401 – 404

Social Network Analysis and Data Mining

Saima Jan¹, Rahila Ruby², Peer Taha Najeeb³, Mudasir Ahmed Muttoo⁴

¹Student, Computer Science Engineering, Kashmir University, Kashmir, India, sammsaima99@gmail.com

²Student, Computer Science Engineering, Kashmir University, Kashmir, India, mlkrahilaruby4@gmail.com

³Student, Computer Science Engineering, Kashmir University, Kashmir, India, peertaha08@gmail.com

⁴Assistant Professor, Computer Science Engineering, Kashmir, India, mudasir.muttoo@gmail.com

Abstract: Data mining is a great innovative technology which helps corporations to focus on the most important information in the data of their stockrooms. Data mining is used in various statistical, machine learning and in graphical methods. Data mining separates the knowledge in to a form which is very useful for many real world applications. Social network analysis has become a very popular field of modern research because it is highly useful for many applications. In this paper we aim to present a review on data mining used for social network analysis.

1. INTRODUCTION

Social network is used to define web-based services that allow individuals to generate a public/semi-public profile within a domain such that they can connect with other users within the network [1]. Social network analysis (SNA) is defined as the study of social networks in order to understand social networks structure and behaviour. Social Networks have received the public imagination in recent years as a proof in the number of popular science and treatment of the subject. Social networks are important sources of online interactions and contents sharing, subjectivity [2], assessments [3], approaches [4], evaluation [5], influences [6], observations [7], feelings [8], opinions and sentiments expressions [9] borne out in text, discussions, reviews, blogs, remarks, news, reactions, or some other documents [10]. Data mining is also a great innovative technology and is a very helpful tool that can help to find different patterns and relationships within the data. Data mining generates hidden information from large databases [11]. Data mining techniques can be used for building descriptive and predictive models of social interactions.

2. SOCIAL NETWORKS AND SOCIAL NETWORK ANALYSIS

A social network is a heterogeneous and multi relational dataset represented by a graph. Vertexes represent the objects (entities), edges represent the links (relationships or interaction) and both objects and links may have attributes [12]. Social networks are usually very large and can be used to represents many real-world phenomena, such as electrical power grids, spread of computer virus and phone calls. Network construction from general, real-world data presents various unexpected challenges owing to the data domains themselves (e.g., information extraction and preprocessing) and to the data structures used for knowledge representation and storage [13].

A social network can be generically understood to be a kind of computer application which facilitates the creation or definition of social relations among the people based on general interests, acquaintance, activities, professional interests,

associative relations, family, and so on. Social networks can arise from information in the sources such as texts, databases, communication systems, sensor networks, and social media. Finding and representing a social network from a data source can be a very difficult. This challenge is due to many factors, including the ambiguity of human language, incompatible representations of information, multiple aliases for the same user, and the ambiguity of relationships between individuals.

2.1 Computer networks as social networks

Computer networks are inherently social networks linking organizations, people, and knowledge [14]. The Data sources include e-mail messages; social networks like Orkut and Yahoo groups; newsgroups like USENET; instant messenger logs like AIM; weblogs like Blogger; and online gaming communities. Computer Science has created the uber-cyber infrastructure for Social Interaction, Knowledge Exchange and discovery. It has an ability to capture difference about various types of social interactions at a very fine granularity with practically no bias.

3. DATA MINING

Data mining is an interactive process within which progress is defined by discovery through either automatic or manual methods. Businesses can learn from their transaction data about the behavior of the customers and thus can improve their business by exploiting this knowledge Web usage information can be analyzed and exploited to optimize information access. Thus data mining generates novel, unsuspected, unbiased, interpretations of data. The main idea of data mining falls under 2 categories [15]:

- a. Predictive data mining: creates the model of the system from the given data.
- b. Descriptive data mining: which generates significant data sets from the existing data.

The aim of these above ideas is achieved by the following data mining techniques [11].

- **Characterization**

Characterization is used to generalize, summarize and possibly different data characteristics.

- **Classification**

Data classification is a process in which the given data is classified in to different classes.

- **Regression**

This process is similar to classification. the major difference is that the object to be predicted is continuous rather than discrete.

- **Association**

It discovers the association between various data bases and the association between the attributes of single database.

- **Clustering**

Clustering involves grouping of data into several new classes such that it describes the data. It breaks large data set into smaller groups to make the designing and implementation process to be simple. The task of Clustering is to maximize the similarity between the objects of classes and to reduce the similarity between the classes.

- **Change Detection**

This method identifies the significant changes in the data from the previously measured values.

- **Deviation Detection**

Deviation detection focuses on the major deviations between the actual values of the objects and its expected values. This method finds out the deviation according to the time as well the deviation among different subsets of data.

- **Link Analysis**

It traces the connections between the objects to develop models based on the patterns in the relationships by applying graph theory techniques.

- **Sequential Pattern Mining**

This method involves the discovery of the frequently occurring patterns in the data.

4. SOCIAL NETWORK ANALYSIS AND DATA MINING

Data mining tools can answer industry questions that traditionally were too time consuming to resolve. Data mining of social networks can be done using the graph mining methods such as classification/topologies, prediction, detection, efficiency, pattern measurement and metrics, modeling, data processing, evolution and structure, and communities [11]. To extract the information represented in graphs one needs to define metrics that describes the global structure of graphs, find the community structure of the network, and define metrics that describe the patterns of local interaction in the graphs, develop efficient algorithms for mining data on networks, and understand the model of generation of graphs.

Social network and its analysis is an important field and it is widely spread among many young researchers. Social networks research emerged from sociology, psychology, statistics and graph theory. Based on theoretical graph concepts, a social network interprets the social relationships of individuals as points and their relationships as the lines connecting them [16]. The various types of social network analysis are;

4.1. 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

4.2. 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.

5. CONCLUSION

The rise of social networks gives very strong effects to the set of techniques developed for mining graphs and social networks. Social networks are rooted in many sources of data and at many different scales. Data Mining provides proficient way to execute and make use of database.

REFERENCES

- [1] Chen, Z. S., Kalashnikov, D. V. and Mehrotra, S (2009). Exploiting context analysis for combining multiple entity resolution systems. In Proceedings of the 2009 ACM International Conference on Management of Data (SIGMOD).
- [2] Asur, S., and Huberman, B (2010) Predicting the future with social network." *Web Intelligence and Intelligent Agent Technology (WIAT), 2010 IEEE/WIC/ACM International Conference on*. Vol. 1. IEEE.
- [3] Kim, Y., Hsu, S.H., de Zuniga, H.G (2013). Influence of social network use on discussion network heterogeneity and civic engagement: The moderating role of personality traits. *Journal of Communication* 63.3, 498-516.
- [4] Korda, H., and Itani, Z. (2013). Harnessing social network for health promotion and behaviour change. *Health promotion practice*, 14(1), 15-23.
- [5] Kaur, G (2013). Social network evaluation criteria and influence on consumption behaviour of the youth segment.
- [6] Bakshy, E., Hofman, J. M., Mason, W. A. and Watts, D. J (2011) Identifying influencers on twitter. In *Fourth ACM International Conference on Web Search and Data Mining (WSDM)*.

- [7] Chou, W. Y. S., Hunt, Y. M., Beckjord, E. B., Moser, R. P. and Hesse, B. W. (2009) Social media use in the United States: implications for health communication. *Journal of medical Internet research*, 11(4).
- [8] Kaplan, A.M. and Haenlein, M. (2010). Users of the world unite! The challenges and opportunities of social media. *Science direct*, 53, 59-68.
- [9] Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis; Foundations and trends in information Retrieval; Vol. 2, 1-2.
- [10] Liu, B (2011). Sentiment analysis and opinion Mining. AAAI-2011, San Francisco, USA, 2011.
- [11] Han J, Kamber M. (2006). Data Mining: Concept and Techniques, 2nd Edition. Morgan Kauffmann.
- [12] Xu G, Zhang Y, Li L. (2011). Web Mining and Social Networking Techniques and applications, 1st edition. Springer.
- [13] Scott J. (2000). Social Network Analysis - A Handbook, Second Edition. SAGE Publication.
- [14] Wellman B (2001). Computer networks as social networks. *Science*. 14;293(5537): 2031-4.
- [15] Du H. (2010). Data Mining Techniques and Applications an Introduction, 1st Edition. Cengage Learning Edition; 2010.
- [16] Mika P. (2007). Social Networks and the Semantic Web, 1st edition. Springer; 2007.

BIOGRAPHIES



Saima Jan is pursuing B.E Degree from SSM College of Engineering & Technology in Computer Science Engineering from University of Kashmir, J&K, India. The field of interests are ASP.Net & networking.



Rahila Ruby is pursuing B.E Degree from SSM College of Engineering & Technology in Computer Science Engineering from University of Kashmir, J&K, India. The field of interests are networking and ASP.Net.



Peer Taha Najeeb is pursuing B.E Degree from SSM College of Engineering & Technology in Computer Science Engineering from University of Kashmir, J&K, India. The field of interests are IOT, Database and Big data.



Mudasir Ahmed Muttoo is the Assistant Professor at SSM College of Engineering & Technology in Department of Computer Science Engineering affiliated to Kashmir University, J&K, India. Area of interests are Computer Networks and ASP.Net.