International Journal of Scientific & Engineering Research, Volume 12, Issue 1, January-2021
ISSN 2229-5518

Exploratory Analysis On The Different Models And Approaches For Social Network Analysis

Mansi Singh, Chandni Deb, Ashna Prasad, Naveen Renji

Abstract- The following paper consists of an overall comparative study of Social Network Analysis. The comparative study provides detailed information and is carried out based on the different models approached for Social Network Analysis. Social network analysis is a method capable of monitoring the interactions in an online environment.SNA can be used to reveal important information about the sentiments of different actors, such as the general activity and the active groups.

Index Terms- Data Analysis, Graph structures, MapReduce, Ontology, Semantic Model, Social Network Analysis Models (SNA), Text mining,

1. Introduction

Social Network Analysis, is the technique of gathering and analysing data from various different social media platforms such as Facebook, Instagram, LinkedIn and Twitter. This technique is commonly used by marketers to track online conversations about products and companies. This analysis process is done with the use of networks and graph theory. It follows the pattern of charatersing the networked structures in terms of nodes and the ties,edges or links that connect them.Social network analysis in general studies the behavior of the individual at the micro level, the pattern of relationships (network structure) at the macro level, and the interactions between the two. Social networks are both the cause of and the result of individual behavior.

To be able to successfully predict a person's next move and thinking SNA is required as it is a step by step process of collecting, training and testing through the different SNA models.Social Network Anaylsis is of primary importance for the whole community to understand the semantics of people coming from different regions, backgrounds and cultures.

2. COMPARATIVE ANALYSIS

A) Different Aspects of SNA

The first paper under consideration proposes three models that are as described below:
The first model is based on **graph-like structure** that is used to describe the network under question. It consists of a display that consists of nodes that represent the actors and edges that represent the relations. Sociologists borrowed this methodology from mathematicians, they renamed this graph-like entity as "sociograms".

The second model proposes the use of **matrices** to represent social network relations. It is rows and columns as there are actors in the data set, and where the entries of this matrix would represent the relations between the actors. The presence of a 0 in the matrix implied the absence of any relation and that of 1, implied that a relationship existed between the two actors (denoted by the row and column of the matrix). It is referred to as an adjacency matrix because it represents who is next to whom in the social space mapped by the relations that we have measured

	Bob	Carol	Ted	Alice
Bob	–	1	0	0
Carol	1	–	1	0
Ted	1	1	–	1
Alice	0	0	1	–

Figure 1.2 Adjacency matrix representation

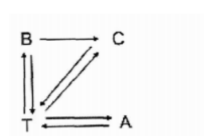

Figure 1.1 A sociogram

The third model revolves around the use of **statistics** to model the probabilities of relational ties between the actors. With the assumption that there are 'n' actors and taking into consideration the information about binary relations between them, it is possible to analyse the data.

The vital properties of the network include maximum flow (the notion how totally connected two actors are, considering the

International Journal of Scientific & Engineering Research, Volume 12, Issue 1, January-2021
ISSN 2229-5518

number of different actors in the neighborhood of a source that would lead to pathways to a target), Hubbell and Katz cohesion (the count of the total connections between actors where each connection is given a weight, according to its length. The greater the length, the weaker the connection) and Centrality and Power (further described by the degree, which refers to the number of ties for an actor, and the closeness, which is the length of paths to other actors).

B) Semantic Model for Academic SNA

The second paper under consideration proposes a novel model called the Semantic Relationship Model (SRM). SRM is an object-oriented model where classes are partitioned into two categories on the basis of their functionality: object classes and role relationship classes. Object classes are user-defined classes that describe the static aspects of real-world entities. Role relationship classes are induced from role relationships and describe the dynamic and many-faceted aspects of real world entities. The model supports five kinds of relationships: regular, contain, role, context, and context-dependent relationships. We distinguish two kinds of attributes: regular attributes and context-dependent attributes. All the information regarding a real-world object is grouped in one instance instead of scattered in several instances. Thus, a real-world entity is represented by one and only one object which can belong to several role relationship classes.

The query language proposed is logic-based. It strictly separates the retrieval part from the result part. In the retrieval part, logical variables are used and they start with $ to get all the values based on their positions whereas in the result part, the manner of construction of the result using the variables bound to various values is defined. The information about objects is organized into hierarchical or composite hierarchies. Thus, path expressions are also introduced.

C) Multidisciplinary Model of dynamic and semantic SNA

The third paper proposes a multidisciplinary model for SNA and is based on cognitive sciences and physics. The model includes two new measures and a predictive system: (1) a measure of tension in the social network, (2) a measure of reactance of a social network used to evaluate the individual stress of its members and (3) an electrodynamic and predictive system for semantic recommendations about social graphs evolutions. The three approaches used in the paper are:

Dynamic graph mining, which may be structure-based where the network propagation phenomenons are studied from a structural viewpoint in several domains such as information technologies, health or environment, or flow-based where the use of physics models is incorporated for static analysis of flows.

According to the **Radio-Electric Principles of Enterprises and Institutions SNA,** some physics models use graphs for the understanding and discovery of theoretical principles on the assumption that the percentages of read, written or shared common documents (e.g. office, mails, instantaneous messages), exchanged data packets (ToIp, VoIp) or other numerical communication marks between individuals, are transposed into electrical intensities, tensions, and powers.

Electrodynamic and Semantic SNA studies the conceptual aspects of a social graph. Based on the principles underlying conceptual graphs theory and semantic networks theory, it usually refers to Semantic Web, Ontology Engineering and logical inferences, with respect to cognitive sciences. Semantic SNA can notably bring real advantages in the areas related to social and human capital management or optimisation of work-groups and working methods, within professional organisations.

D) Content Based Social Network Analysis Model

In this paper, a Novel model for social Network Analysis is presented which analyses their communicative content rather than analysing their quantity of relationships.The Content based model uses Text mining and Clustering techniques to capture the content of communication and to identify the most popular themes.

Content Based Social Network Analysis with the aid of deep linguistic analysis is done in three steps:

1.Concept Analysis

The objective of this phase is to identify the emergent semantics of a community, i.e. the concepts that better characterize the content of actor's communications. Concepts are extracted from available texts (hereafter referred to as the domain corpus) exchanged among the members of the community.

2. Topic Detection

International Journal of Scientific & Engineering Research, Volume 12, Issue 1, January-2021
ISSN 2229-5518

A topic is a set of semantically close concepts, but the relevance of a topic is tied to the specific set of documents that characterize interactors' communications in a given time interval.It is a clustering task.The objective is to organize concepts in groups, or clusters, so that concepts within a group are more similar to each other than are concepts belonging to different clusters.

3.Social Network Analysis

The Content Based- Social Network is then modelled through an undirected graph with:

- the nodes representing the groups *gi*;
- the edges representing the similarity between nodes measured by the *cosine* function:

$$\cos\text{-}sim(g_i, g_j) = \cos(g_i, g_j) = \frac{I_{g_i} \bullet I_{g_j}}{\left|I_{g_i}\right|\left|I_{g_j}\right|}$$

Social Network Measures:

To analyze the network, we selected the following network analysis measures:

- Average Degree Centrality:

$$ADC = 1/(N(N-1))\sum_{i=1}^{N}\deg(i)$$

where *deg(i)* is the number of edges connected to a node *i* and *N* is the number of nodes in the network It measures whether the network is weakly or strongly connected.

- Degree Centrality of a Vertex:

$$DC(v) = \deg(v)$$

It measures the connectivity of each social actor.

- Weighted Degree Centrality of a Vertex:

$$DC_w(v) = \sum_{e \in P_v} w(e)$$

where *Pv* is the set of edges connected to the node and *w(e)* is the weight of the edge *e*. It measures connectivity of each social actor by taking into account the edge's weight.

E) MapReduce Based Social Network Analysis

In the fifth paper, a large-scale Social Network Analysis is done based on MapReduce.In this paper, Hadoop an open source implementation of Mapreduce has been used to conduct a series of analysis on social networks including several distributions,clustering coefficient and diameter.Google had proposed a programming paradigm called MapReduce which is capable of processing large-scale data with inexpensive machines.

MapReduce is used to measure the degree distribution, the node strength distribution, the
edge weight distribution, the diameter and the clustering coefficient.

A "user posting" network dataset from Delicious and a coauthoring social network by the DBLP was taken and distributed to 20 computers and used MapReduce programming model to get the correlation networks.Two correlation networks were obtained which are undirected weighted graphs.MapReduce methods are used to capture the structure and characteristics of the two social networks.The experiments were carried out on a Hadoop cluster made up of 20 machines.

Results that were obtained after carrying out experiments on a Hadoop cluster.

1. *Degree Distribution*

The degree distribution of social networks conforms to power laws.

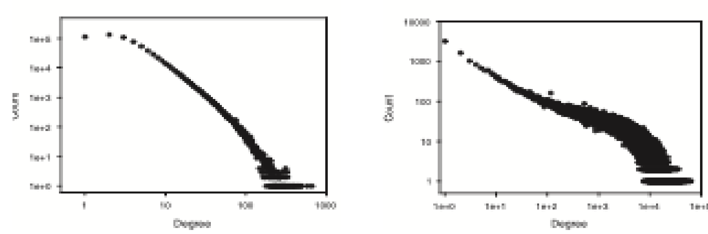

Figure 2 Result of Degree Distribution

2.Weight Distribution

International Journal of Scientific & Engineering Research, Volume 12, Issue 1, January-2021
ISSN 2229-5518

A network's weight distribution obeys power law.
Figure below shows the parameter of power law.

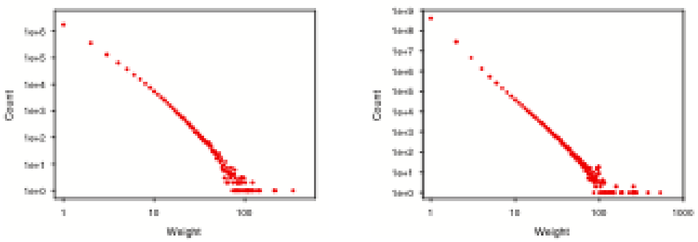

Figure 2.1 Result of Weight Distribution

3.Node strength Distribution

The strength of a node indicates the information about it's connectivity and weights of it's links.
It is observed that the node strength distribution of a co-relationship network obeys power law,too.

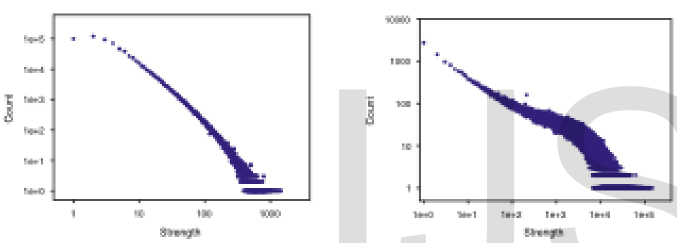

Figure 2.2 Result of Node Strength

4.Clustering Coefficient
After calculating the mean clustering coefficients of all nodes, it was found that the Delicious dataset was too large and that of DBLP dataset is 0.581.The clustering coefficient of real world is much bigger than random networks.

5.Diameter
In DBLP dataset, there were 10% nodes that were isolated and it's diameter is 14 and that of Delicious dataset is 11.

A new way of analysing the structural properties of two social networks using MapReduce has been presented.

F) Do You Know the Way to SNA?: A Process Model for Analyzing and Visualizing Social Media Data

In the sixth paper, the **Network Analysis and Visualization (NAV) process model** is derived. It emerged from the collective experience of students learning to use SNA metrics and visualizations. The stages within the model where interventions from peers, experts, and analysis tools are most useful were identified.

The analysis began with the compilation of the data into individual student profiles containing survey responses, diaries, observation notes and interview transcripts, assignments with peer and instructor comments, and grades. Based on initial

observations, a list of open-ended questions was created that was intended to be answered from the study data. These were elaborated on and modified as summary reports came to be created on selected issues. Such reports became the basis for in-depth analysis to identify common themes and patterns across students.

The SNA tool used was **NodeXL,** which is a plug-in for Excel 2007 that exploits a widely used spreadsheet paradigm to provide a range of basic network analysis and visualization features.

G) Ontology-Based Text-Mining Model For Social Network Analysis

Our seventh paper discusses a system model that is applicable for analyzing the unstructured data inside social media posts on electronic products. For the analysis, posts on social networking websites have been mined and the keywords are extracted from such posts. The extracted keywords and the ontologies of electronic products and emotions form the base for the text mining model which is used to understand online consumer behavior in the market.

First and foremost there is a need to understand the meaning of "Ontology". Ontology is defined as a set of representational primitives used to model a domain of knowledge. It is a formal and declarative representation, which includes the
vocabulary (or names) for referring to the terms in a specific subject area and the logical statements that describe what different terms are and how they are related to each other.

This paper discusses an ontology-based text mining model in detail. By using the ontology modeling environment,an ontology primarily based on electronic products has been created. Additionally, the ontology based on emotions and language used on social networking sites is also developed and merged with the electronic products ontology to analyze consumer behavior on social networking sites.

Emotions Ontology : the emotions ontology is essentially used for describing the emotions of consumers codifying their feelings related to these different products. The ontology starts with categorizing sentiments, which has the subclasses happiness and sadness. The keywords under the subclass

International Journal of Scientific & Engineering Research, Volume 12, Issue 1, January-2021
ISSN 2229-5518

happiness are; cool, great, enjoy, fun, eager, smiling, excited, glad, joy, love, happy, good, excellent, pleased, etc. The keywords associated with the subclass sadness are: bored, tired, frustrated, disliked, sad, disappointed, bad, worst, etc.

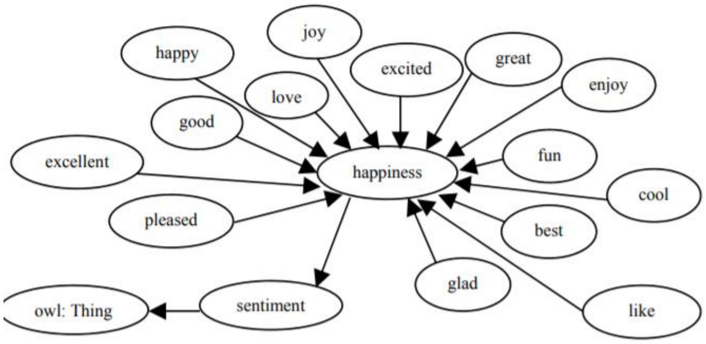

Figure 3 Illustrating the emotions ontology

Electronics Product Ontology: The ontology aims to provide a controlled vocabulary to semantically describe concepts about electronic products and it also demonstrates the relationships between the various concepts. The ontology starts with the element type of product: computer, computer-related products and household products. Then, their respective subcategories are built and the product attributes are summed up as the name, variety, type, size, and price used to describe all of the sub-categories.

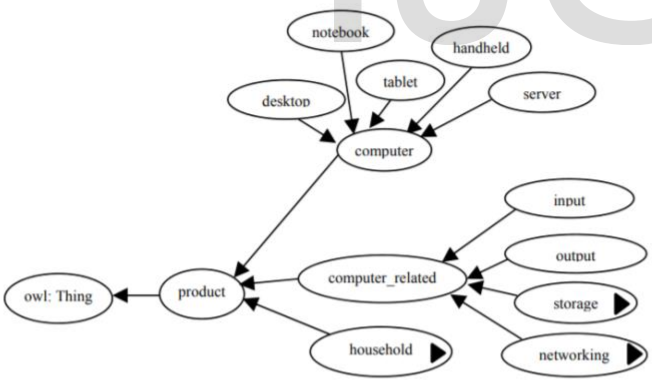

Figure 3.1 Illustrating the electronics product ontology

Ontology based Text Mining Model for SNA: The text model for text mining consists of four major modules.

1. Ontology management module
2. User query processing moule
3. Information foundation module
4. Query Analysis Engine Module

3. CONCLUSION

In this paper, we've reviewed Social Network Analysis using various models and social networks properties. Social network analysis provides some useful tools for addressing many aspects of social structure.

MapReduce programming model shows it is the most powerful in processing large-scale social network data and can replace many traditional analysis methods.

An innovative system defined in the context of a multidisciplinary approach that enables predictions and recommendations on the evolutions of a social network. This work is a baseline for the development of new decision-making functions and tools, for socio-professional troubles risk prevention, performance loss risk prevention and social risk prevention in enterprises and institutions.
The data is mined from a frequently updated source, for example posts in the case of a social networking site. Therefore a system that updates associations between entities based on these additions is desirable. The extraction of relationships based on new content can modify earlier dependencies, yielding better results. By using a query analysis engine that maps relations to ontology; the system has the potential to display better semantic search capabilities.

The language used on social networking sites is non-rule based. It does not follow the syntactical rules of the English language. The development of a branched and detailed ontology is required.

Social network analysis finds its applications in many fields. With graph theory as its foundation it has become a multidisciplinary approach with applications in sociology, the information sciences, computer sciences, geography etc

References

[1]Wasserman, S. & Faust, K. Social Network Analysis. Cambridge University Press (1994).
[2] A. Hanneman and M. Riddle, "Introduction to social network methods," online.http://www.faculty.ucr.edu/hanneman/nettext/, 2005.

[3] A. Swartz and J. Hendler, "The semantic web: A network of content for the digital city," in Proceedings of Second Annual Digital Cities Workshop, Kyoto, Japan, October 2001.
[4] J. Hendler, T. Berners-Lee, and E. Miller, "Integrating applications on the semantic web," Journal of the Institute of

International Journal of Scientific & Engineering Research, Volume 12, Issue 1, January-2021
ISSN 2229-5518

Electrical Engineers of Japan. vol. 122. no. 10. pp. 676-680, 2002.

[5] L. Freeman, Graphic techniques for exploring social network data. New York: Cambridge University Press. 2005, pages 248-269 in 'Models and methods in social network analysis' edited by P.J. Carrington, J. Scott, and S. Wasserman.

[6]S. Bhagat, G. Cormode, and S. Muthukrishnan. Node classification in social networks. In *Social Network Data Analytics*, pages 115–148. 2011.

[7]G. Costa and R. Ortale. A bayesian hierarchical approach for exploratory analysis of communities and roles in social networks. In *ASONAM*, pages 194–201, 2012.

[8]S. Das, O. Egecioglu, and A. E. Abbadi. Anoṅimos: An lp-based approach for anonymizing weighted social network graphs. *IEEE Trans. Knowl. Data Eng.*, 24(4):590–604, 2012.

[9] R. He and T. Zheng. Estimation of exponential random graph models for large social networks via graph limits. In *ASONAM*, pages 248–255, 2013.

[10]N. F. Noy and D. L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology", Stanford University, USA, 2001.

[11] J. Lin. Graph algorithms with mapreduce. In *http://www.umiacs.umd.edu/jimmylin/cloud-2008-Spring/Session4.ppt*.

[12] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *Proceedings of ACM IMC07*, pages 29–42. ACM, 2007.