International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:268
Vol.2, No.12, December 2013 DOI:10.15693/ijaist/2013.v2i12.30-32

Network Analysis of Ontologies: A step towards Knowledge Discovery in Social Networks

Dr. Jyoti,
Asst.Prof.(CE),
YMCA University of Science and Technology,
Faridabad, Haryana, India

Abstract: Ontology is a term that appears in contexts as diverse as computer science, linguistics and philosophy. In computer science, ontology formally represents knowledge as a set of concepts within a domain, using a shared vocabulary to denote the types, properties and interrelationships of those concepts. Social networks are explicit representations of the relationships between individuals and groups in a community. This paper aims to provide the insight about the different work done in the area of knowledge discovery in social networks based on ontologies.

Index Terms: Ontologies, Social Networks, Knowledge Discovery

I. INTRODUCTION

Knowledge Discovery can be defined as a process which aims at the extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) information from data in large databases [3]. Knowledge discovery approaches adopt the methods developed in Machine Learning and Data Mining [4, 5 and 6] which provides techniques for data analysis with varying knowledge representations and large amounts of data, and also methods developed in statistical learning [7] and pattern recognition [8] contributing data analysis in general.

In abstract, social networks are simple graphs with nodes for the people and groups and links for the relationships. In practice, the links can encode all kinds of relationships which may be familial, friendship, professional or organizational. Social network theory, the study of such social networks, has developed techniques found useful in many fields, including sociology, anthropology, psychology and organizational studies. The knowledge thus discovered could handle different issues which may pertain to developing semantic web or tackling issues of security on social networks and many more.

A key argument for modeling knowledge in ontologies is the easy reuse and re-engineering of the knowledge. However, besides consistency checking, current ontology engineering tools provide only basic functionalities for analyzing ontologies. Since ontologies can be considered as labeled and directed graphs, graph analysis techniques are a promising tool. In late 1970's, Social Network Analysis emerged as a major research area. Its aim is to analyze the structure of the social communities. Typical applications include the analysis of relationships like friendship, communication patterns (e. g., phone call graphs), and the distribution of attendants over several events. While social structures are currently a steeply rising topic within the Semantic Web community, Social Network Analysis (SNA) has only been applied marginally on ontologies and the Semantic Web. Semantic Network Analysis *(SemNA)* is a sub area of semantic web mining [9] that addresses the mining of semantic web. A systematic development of Semantic Network Analyis (SemNA), is slowly gaining pace as the adoption of SNA to ontologies and the Semantic Web.

II. BACKGROUND

Social network analysis (SNA) is the mapping and measuring of relationships and flows between people, groups, organizations, animals, computers or other information/knowledge processing entities. The nodes in the network are the people and groups while the links show relationships or flows between the nodes. SNA provides both a visual and a mathematical analysis of human relationships [10]. The two aspects of SNA, the functional aspect and the structural aspect, each highlight a different perspective of research. The functional view focuses on how the function of a network is determined by the structure of a given network. Thus the question of flow between nodes is very prominent. The structural view on the other hand is more interested in the question of structure per se and what statements about a given network can be made based on the analysis of structure alone. Both aspects can be viewed separately, but for some objects of interest, such as organizations, a combined approach may be more appropriate. Since the use of SNA tools in the semantic web environment is just starting out, we will focus in this paper on the structuralist view on SNA, in particular on different notions of centrality. The concept of centrality has many different branches. Just to name a few: in/out degree centrality, betweenness centrality, information centrality, eigenvector centrality. For a good overview see [11].

Authors of [12] describe to a great extent the history of rank prestige index, which is an eigenvector centrality based concept. This index is based on the idea that the rank of a group member depends on the rank of the members he or she is connected to. Stated in mathematical terms this yields the eigen value equation (for an eigen value equal to 1). The components of the principle eigen vactor are the rank prestige indices of each group member. This component is implemented in the hub-and-authority-algorithms of Kleinberg [13] and also in pagerank algorithm proposed by Page and Brin[14].

There have been different approaches to the analysis of unbalanced graphs. All concepts work very well on undirected and un-weighted graphs. But if none of these restrictions apply for a given graph, difficulties arise. Freeman [15] proposed to use the possibility to split any asymmetric square matrix into its symmetric and skew-symmetric part, perform a singular value decomposition of the skew-symmetric matrix, and showed, that the result could be interpreted as a ranking of dominance. Authors of [16] could identify subgroups in unbalanced email networks by analyzing betweenness centrality in the form of intercommunity edges with a large betweenness value. These edges are then removed until the graph decomposes into separate communities, thus re-organizing the graph structure. Barnett and Rice [17] showed that the transformation of asymmetrical data into matrices that avoid negative eigenvalues may result in the loss of information. This is why authors of [18] have transformed the adjacency matrix into Hermition matrix

A. *Prior Work on Network analysis of ontologies*

Hoser et al. [8] performed social network analysis on the SUMO (Suggested Upper Merged Ontology) and SWRC (Semantic Web for Research Communities) ontologies. They found that social network analysis provide useful insights into the structure of ontologies. They found the need to preprocess ontologies to a simpler structure prior to the social network analysis. In this paper the authors explored the use of centrality analysis on the ontologies. They specifically identified betweenness centrality and eigenvector centrality for these two ontologies. The authors consider the betweenness centrality useful in identifying the core concepts in the ontology.

Stuckenschmidt [19] analyzed ontologies and used relative strengths to determine if ontology needs to be partitioned. In the paper the author represented the ontology as a proportional strength network where the weight of the relationship is determine by the inverse of the degree of the node. The partitions were then determined by applying minimal cut algorithm on the graph.

Coskun et al. [20] used social network analysis on ontologies to identify concept groups. In this paper the authors investigate nine different representation of ontology as a graph. The three basic representation being a plain RDF graph structure, a graph where the predicates are also represented as nodes and a third where only the classes are represented as nodes. Each of these representations had two extensions, one where the literals were ignored and another where the RDF, RDFS, OWL and XML Schema nodes were ignored.

Social network analysis has also been used for the development of ontologies [21].

B. *Analysis of ontologies*

- RDFS Schema
  This class of ontologies is described using the RDFS language. These tend to be smaller and more basic than the OWL based ontologies.

  1. FOAF (Friend of a Friend) is ontology to describe the details of a person. [22]
  2. WOT (Web of Trust) is an ontology to facilitate signing RDF documents. [23]
  3. DOAP (Description of a Project) is an ontology to describe software projects [24]
  4. Dublin Core is an upper ontology from the Dublin Core Metadata Initiative [25]
  5. AtomOwl is the ontology behind the Atom syndication format. [26]

- OWL Ontologies
  This class of ontologies is described using the OWL language. They are generally much richer and use the more advanced concepts provided by OWL. The number of statements in OWL ontologies is usually an order of magnitude higher than those in RDFS Schemas.

1. SWEET (Semantic Web for Earth and Environmental Terminology) is ontology for environmental terms [27].
2. COSMO (Common Semantic Model) is a foundational ontology containing basic and primitive concepts [19].
3. OpenGALEN is an ontology to represent clinical information. The Common Reference Model (CRM) is the core of the ontology; the Diseases Extension is one of the sub ontologies. Version 8 of the ontology was used for this analysis [20].
4. SUMO (Suggested Upper Merged Ontology) is an upper ontology for general purpose terms [21].

III. CONCLUSION

Social Network Analysis provides promising set of tools for analyzing ontologies and semantic web applications, providing depp insights into the structure of ontologies and knowledge bases. Analysis of ontology can be done at different levels of granularity. The gained insights may help to design or redesign ontologies in order to find redundancies. This paper provides an insight of the work of different authors in this regard.

REFERENCES

[1]Gruber, Thomas R., “, Knowledge Acquisition”, Vol 5 issue 2, 199–220, June 1993.
[2] Arvidsson, F., Flycht-Eriksson, A. “Ontologies I”, Retrieved 26 November 2008.
[3] Fayyad U, Piatetski-Shapiro G, Smith P and Uthurusamy R (eds.) (1996) Advances in Knowledge Discovery and Data Mining. MIT Press, Cambridge, MA.
[4] Mitchell, Machine Learning. McGraw-Hill, 1997
[5] Witten I H, Frank E, “Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations”, Morgan Kaufmann, San Francisco, 1999.
[6] Hand DJ, Mannila H. Smyth P, “Principles of Data Mining (Adaptive Computation and Machine Learning)”, MIT Press, Cambridge, MA, 2001.
[7] Hastie T,Tibshirani R and Friedman J H, “The Elements of Statistical Learning: Data Mining, Inference, and Prediction”, Springer Series in Statistics, Springer, New York, 2001.
[8] Duda R O, Hart P E and Stork D G, “Pattern Classification”, 2nd edition, Wiley, New York,2000.
[9] George A. Barnett and Ronald E. Rice, “Longitudinal non-euclidean networks: Applying Galileo, Social Networks”, 7:287–322, 1985.
[10] Orgnet.com, Social network analysis software and services for organizations and their consultants http://www.orgnet.com.
[11] M.G. Everett and S.P. Borgatti, “The centrality of groups and classes. Journal of Mathematical Sociology”, 23(3):181–201, 1999.
[12] Stanley Wasserman and Katherine Faust. Social Network Analysis: Methods and Applications, volume 8 of Structural Analysis in the Social Sciences. Cambridge University Press, Cambridge, 1 edition, 1999.
[13] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment, in Ninth Annual ACM-SIAM symposium, 668-677, 1998.
[14] Sergey Brin and Lawrence Page. The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer Networks and ISDN Systems, 30(1-7):107–117, April 1998.
[15] Linton C. Freeman, “Uncovering organizational hierarchies”, Computational & Mathematical Organization Theory, 3(1):5 – 18, 1997.
[16] Joshua R. Tyler, Dennis M. Wilkinson, and Bernardo A. Huberman., “Email as spectroscopy: Automated discovery of community structure within organizations”. cond-mat/0303264, 2003.
[17] George A., Barnett and Ronald E. Rice. , “Longitudinal non-euclidean networks: Applying galileo., Social Networks, 7:287–322, 1985.
[18] Hoser, B et al., “Semantic Network Analysis of Ontologies”, Proceedings of the 3rd European Semantic Web Conference, 2006.
[19] Heiner Stuckenschmidt, “Network Analysis as a Basis for Partitioning Class Hierarchies”,In Proc. of the ISWC2005 Workshop on Semantic Network Analysis, 2005.
[20] G. Coskum, M. Rothe, K. Teymourian, A. Paschke, “Applying Community Detection Algorithms on Ontologies for Identifying Concept Groups”, In O. Kutz and T. Schneider (Eds.) Modular Ontologies, IOS Press, 2011.
[21] Peter Mika, “Ontologies are Us: A Unified Model of Social Networks and Semantics,” In Yolanda Gil, Enrico Motta, V. Richard Benjamins, and Mark A. Musen, editors, ISWC 2005, volume 3729 of LNCS, pages 522–536, Berlin Heidelberg, November 2005.
[22] Stuckenschmidt, H. & Klein, M., “Structure based partitioning of large concept hierarchies,” In Proc 3rd International Semantic Web Conference ISWC, 2004.
[23] Web of Trust RDF Ontology. http://xmlns.com/wot/0.1/
[24] Description of Project (DOAP), 2008, http://trac.usefulinc.com/doap
[25] DCMI Usage Board. DCMI Metadata Terms.2010. http://dublincore.org/documents/dcmiHterms/
[26] AtomOwl. 2006. http://bblfish.net/work/atom-owl/
[27] Semantic Web for Earth and Environmental Terminology (SWEET). 2011, http://sweet.jpl.nasa.gov/ontology/
[28] Common Semantic Model (COSMO). 2009 http://ontolog.cim3.net/cgiHbin/wiki.pl/COSMO /
[20] Rector, A.L. et al., “OpenGALEN: open source medical terminology and tools”, in the Proc. of AMIA Annual Symposium proceedings AMIA Symposium AMIA Symposium, 2003.
[21] Niles, I. & Pease, A., “Towards a standard upper ontology”, C. Welty & B. Smith, eds. Proceedings of the international conference on Formal Ontology in Information Systems FOIS 01, p.2H9, 2001.

AUTHOR PROFILE

Dr Jyoti is an Assistant Professor (Computer Engineering) at the YMCA University of Science & Technology, Faridabad. She has about 11 years of experience and earned her Doctorate (Ph.D) from the above university in 2011. She has published a number of papers in International journals and her areas of interest and research include: web mining, web pre-fetching and object oriented programming among others. Dr Jyoti is currently also guiding research scholars for their respective Doctorates.