International Journal of
Advanced Research in Social Sciences, Environmental Studies & Technology
Hard Print: 2536-6505 Online: 2536-6513

September, 2016

Vol. 2, No. 1

Review Social Network Analysis and Mining: Link Prediction

[1]Kabir Ismail Umar &
[2]Fatima Chiroma
[1]Department of Information Technology,
Modibbo Adama University of Technology Yola
[2]Software Department,
American University of Nigeria

Abstract

The rapid generation and uncontrollable accumulation of the social network data has raised a real issue now, because the data are vast, noisy, distributed, unstructured and dynamic. Since this data can be mined by using web mining techniques, social network analysis and link prediction algorithms, this article tries to understand the social structure and issues surrounding mining social network data. We will also be looking at the link prediction problems in dynamic social networks and the important techniques that can be applied as an attempt for a resolution.

Keywords:
Social Network, Social Network Analysis, Link Prediction and Web Mining.

Corresponding Author:
Kabir Ismail Umar

Background to the Study

The social network, which is a social structure comprising of people who are linked by different types of interdependency, has changed the nature of information in terms of volume, availability and importance (G. Nandi and A. Das, 2013). According to Gupta et. Al., (2015) a social network is denoted by a graph fig. 1, where the nodes represent a user and edges represent relationships among nodes which are the users.

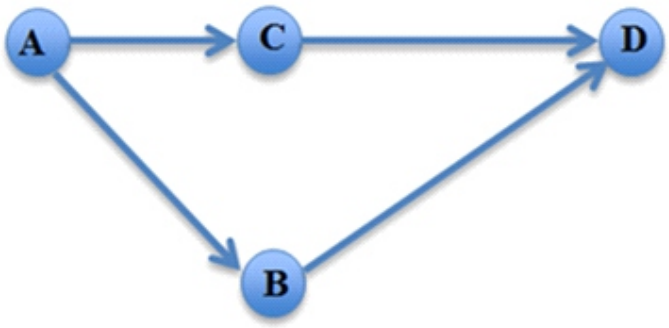

Fig. 1. Social Network.

It is normally formed by continuous interactions between people. Gupta et. al., (2015) claimed that social networks are dynamic in nature, as they grow over time through the addition of new users, creation of new relationships and ending of some old relationships. Nandi et al strongly agrees that social networks are dynamic and even went further to state that the data generated from online social network is vast, noisy and distributed as well therefore to analyze such complex and dynamic social network data appropriate data mining techniques are required.

Through the Social Network, participants publish or publicly reveal a lot of personal information (communications, relationships and behaviors) and this information have real values that can be mined, utilized and monetized. With this information interested parties can examine, determine and even predict things about a user or group of users for many purposes such as to improve decision making, control costs and minimize risks. This data also has research implications that will enable researchers to improve the design and robustness of several systems such as recommender engines. Users on social networks publish personal information that has real economic values. With this information, interested parties can examine and predict things about a user or group of users for many purposes. Such as to improve decision making, control costs and minimize risks. This scenario is showing Fig. 2 as a Link Prediction Causal Loop.

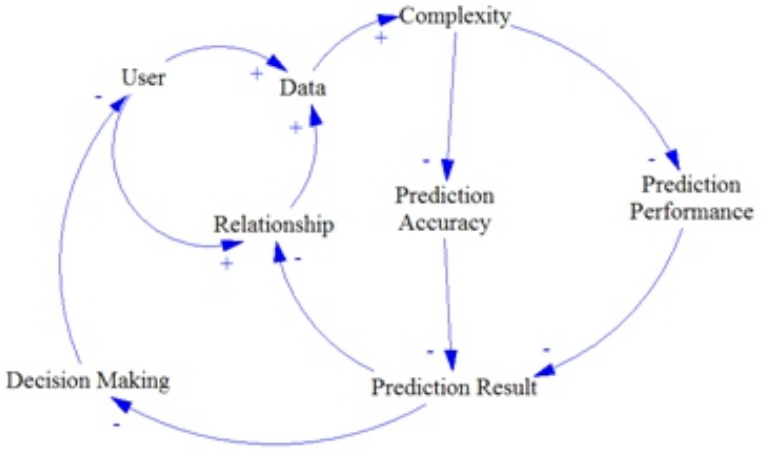

Fig. 2. Link Prediction Causal Loop

Social networks are highly dynamic objects as they grow and change quickly over time through the addition of new edges, signifying the appearance of new interactions in the underlying social structure (Liben-Nowell and Klienberg, 2007). As social networks continue to grow, excess data needs to be mined and Chakrabarti (2003) claimed that web mining is the most suitable. Web mining is an application of data mining (which is the technique of discovering and extracting useful information from large data sets or databases) used to discover and extract useful information from the web" (Hand et. al., 2001). There are three web mining techniques namely the web content mining, web usage mining and web structure mining. These techniques have been used to mine contents, discover interaction patterns between users and identify structures and links between web pages.

Additionally, Ting (2008) stated that social network analysis is a vital technique that will aid the understanding of social behaviors, social relationships and social structure (Nandi, and Das, 2013) . According to G. Nandi, and A. Das (2013) "Social network analysis is the mapping and measuring of relationships and flows between people, groups, computers and other connected information/knowledge entities which provides both a visual and a mathematical analysis of human relationships". It is similar to web mining but it is about data extraction from different resources (Nandi, and Das, 2013) . Other important and popular techniques that are used for social network mining are the data mining techniques such as sequential pattern analysis, visualization, association rule generation, clustering and classification which have been used by several researchers.

Additionally, Borgatti claimed that in the past the real problem with social network analysis (and mining) was its inability to statistically test hypotheses however this is less of a problem now with the advent of permutation tests (Hand et. al., 2001). Although, that is not to say that the social network analysis and mining problem has been eradicated in fact there are still several problems surrounding it and even Nandi et al stated that the problems related to mining social networks are still in the infancy state as such there is subsequent need to develop techniques for further improvements (G. Nandi, and A. Das, 2014). Although, to

specifically mine data to make predictions, link prediction algorithm is additionally required because it is a "link mining task that tries to find new edges within a given graph"and attempt to imitate them (Gastulla, and Cortes, 2014). According to S. Yu, and S. Kak (2012) most social media predictions can be done better by expert human agents however there is need to automate predictions. For example, automated predictions are unbiased, cost effective, more accurate and performs better than human agents.

The problem statement here is given any social network graph at a time t (Fig.3), accurately predict the new relationships that will be created on the network after an interval of time t1.

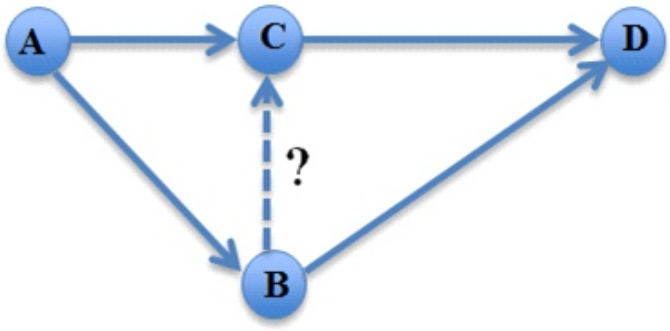

Fig. 3. Link Prediction at time t1.

Literature Review

Research interest in the area of Social Network Analysis specifically link prediction (scalability) is increasing due to its importance. Previous research works have revealed several problems with link prediction such as scalability, accuracy and efficiency/performance. Gupta el. al., (2013) Nandi, and Das (2014) in particular have recently worked on related research and they have emphasized on the importance of finding a solution to the scalability problem of link prediction. Fire et. al., (2011) has also done some research where he claimed that the massive growth of social networks has brought about several research directions in which the structural and behavioral properties of large-scale social networks are examined. Additionally, Gupta et. al., (2015) did some work which agrees with Fire el at claim but they were more specific as they believe that the dynamic change forms the base of link prediction algorithms which involves trying to understand the process of the dynamic changes and trying to replicate them. Additionally, Nandi and Das (2013) have identified seven other key representative research issues in their research namely influence propagation, community or group detection, expert finding, recommender systems, behavior and mood analysis, predicting trust and distrust among individuals, and opinion mining.

Furthermore, in Fire et. al., (2011) they also described the link prediction (problem) as a "problem of predicting the existence of hidden links or creating new ones in social networks". They further stated the relevance of this problem to different circumstances as in recent years several algorithms have been proposed so as to solve the problem but majority of which were merely tested on bibliographic or co-authorship data sets (Fire et. al., 2011). Although not disputing Fire's claim for the description of the link prediction problem,

Nandi and Das (2013) stated that research on link prediction has evolved over the years however the main concern is the scalability of solutions that have were proposed for link predictions. Fire et al did however state that in addition to the "link prediction problem", "existing link prediction techniques lack the scalability required for full application on a continuously growing social network" (Fire et. al., 2011). Gupta et. al.,(2015) similarly agrees that there is a scalability problem with dynamic networks as they earlier stated that link prediction algorithms involves trying to understand the process of the networks dynamic changes and try to replicate them.

In their work, Nandi and Das (2014) concluded that most of the existing work on link prediction focuses on building models that are based solely on the structure of the network, while others were built based on categorical attributes of the nodes but they strongly believe that algorithms and techniques can still be developed that will provide more accuracy and speed that is based on implicit social networks which is formed due to the daily interaction between users i.e. dynamic networks. They further suggested that the key factors to consider while designing a new propagation model is its reliability on very large number of parameters but most importantly it should have the ability to handle the problem of scalability (Nandi and Das, 2014) (Wang et. al., 2015) (Xiaoyi, 2014).

Result/ Analysis

The result of the analysis will provide an informative rather than conclusive result, which will serve as an evidence of the validity of the proposed method. The presentation of the results will be well structured and readable, while the result itself will be vital to readers that plan on using the proposed method as it will give an idea on what to expect. The result can also be used as a benchmark or means of comparison to better comprehend the method. Since this proposed method will be applied to a social network which is known to be dynamic, knowing the size of the data used is very important because the proposed method may not be scalable or applied to certain sized data. Additionally, the logic behind the choice of social network will be presented as it will aid in knowing which type of social network this proposed method can be applied to.

Conclusion

The best approach to study the question is to use the quantitative approach/method to carry out a comparison analysis of the different link prediction algorithms in a social network using performance metrics as a base and similarity as a measure of calculation. Furthermore, the main reason for using the quantitative approach is because a conclusive and more descriptive result is required to arrive at an accurate conclusion as such a standard measurement is used to avoid unfairness. Additionally, the data that will be used from Facebook to analyze the algorithm is structured in the form of numbers so as to get a result that can be used to measure the accuracy of the link prediction algorithms.

Research in this area is in its infancy stage and the interest is increasing due to its importance. Therefore, it is essential for researchers to fine tune the approaches and processes used for Social Network Analysis and Mining especially link prediction, in order to obtain efficient and accurate results. That is the purpose of this article as it will contribute to overcoming the challenges in social network data analysis and mining by attempting to resolve the link prediction problem thru proposing a scalable framework/model based on the analysis carried out. Furthermore, if link prediction is effective and accurate for dynamic

social networks the results of the analysis will be beneficial to other network domains. That is, this research will as well benefit other fields in obtaining maximum accuracy, such as Bioinformatics, E-commerce and Security, it has become critical that link prediction algorithms are accurate on small and large datasets or dynamic networks.

References

Chakrabarti, S. (2003). *Mining the web: discovering knowledge from hypertext data.* California: Morgan Kaufmann publishers.

Cortes, D. G.-G. (2014). Link prediction in very large directed graphs: exploiting hierarchical properties in parallel. *3rd workshop on on knowledge discovery and data mining meets linked open data-11th Extended Semantic Web Conference*, (pp. 1-13).

Das, G. N. (2014, April). Online social network mining: current trents and research issues. *International Journal of Research in Engineering and Technology, 3*, 346-350.

Das, G. N. (n.d.). A survey on using data mining techniques for on-line social network analysis. *International Journal on computer science, 10*, 52-57.

Elovic, M. F. (2011). Link Pridiction on social Network using Computationally Efficient Topological Features. *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust(PASSAT) and 2011 IEEE Third International Conference on social Computing (Socialcom)*, (pp. 73-80).

Ferrara, E. (2012). *Mining and analysis of social networks, PhD.dissertation* . Messina: Department of Mathematics University Press.

Kak, S. Y. (2012). *A survey of prediction using social media.*

Klienberg, D. L.-N. (2007, May). The prediction problem for social networks. *Journal of American society for Information Science and Technology, 58*, 1019-1031.

Lanzi, F. F. (2003). *Recent development in web usage mining research*. LNCS 2737, Ch Data Warehousing and Knowledge Discovery. 140-150.

Shukla, S. G. (2015, February). Comparison analysis of link pridiction algorithms in social network. *International Journal of computer Applications, 111*, 27-29.

Smyth, D. H. (2001). *Principles of data mining.* Massachusetts: MIT Press.

Sravanthi, E. R. (2012, October). Analysis of social network using the techniques of web mining. *International Journal of Avanced Research in Computer Science and Software Engineering, 2*, 443-450.

Ting, I. (2008). Web mining techniques for on-line social network. *Proceedings of International Conference on Service Systems and Service Management*, (pp. 1-5).

Xiaoyi, L. (2014). A deep learning approach to link prediction in dynamic networks . *Preceedinds of the 2014 SIAM International Conference on Data Mining*, (pp. 289-297).

Zhou, P. W. (2015, January). *Link pediction in social networks: the state-of-the-Ar*t. Science China information Sciences. *58*, 1-38.