International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-1, October, 2019

A Correlative Scrutiny for Improving the Career Guidance Links in Social Network

Sudalai Muthu T, Rohini A

Abstract: ***Social Network analysis techniques have shown the ability to meet educators and influence career development guidance. Career education mentoring plays an important role in the career supporting which are the interest, the strength, and the aspirations of the students. In this paper, we proposed a career development framework enhances the node influence propagation and effective interaction between nodes is taken as a strong link, from the node influence propagation we divide it into two categories such as: (1) career predictions to persuade prospective graduates to pursue the desired career path, with career prospects considered by them as a learning opportunity. (2) Social network analysis and persuasive techniques are used to motivate within a social networking framework where there is a tendency to adopt desired professional behaviors. The process begins with the discovery of behavioral features to create a cognitive profile and to identify individual disabilities. We compare a clustering algorithm that predicts the accuracy values and pattern of creations to a social network for achieving collaborative cognition.***

Keywords: Social Network, Community and Link Analysis.

I. INTRODUCTION

The fastest-growing Social Networking Site is Facebook. It's been widely accepted as a platform for communications and collaborations. The students and Higher Education Academicians are used to sharing their valuable resources. It's been notified that the students are far more ahead in using communication technologies as compared to the Higher Education Institutions Academicians. In our research work, the study analyses the social network graphs which are Facebook pages of groups having members like career guidance to the students, etc. The study discovered that in some networks, members are less, but online posts and messages are huge. Student's active participation or interactions in social networks can be analyzed by using Social Network Analysis (SNA) approaches.

In our study, we explore the techniques for the SNA to detect the patterns in the Online Social Media Networks (OSMNs). These patterns are generated by the interactions and the content exchanged among the users and students. The latest platforms which have the potential to make everyone physically disconnected, but logically connected in the social networking sites.

Revised Manuscript Received on October 30, 2019.

* Correspondence Author

SudalaiMuthu T*, Associate Professor, Department of Computer Science and Engineering, Hindustan Institute of Technology and Science, Chennai, India.

Rohini A, Research Scholar, Department of Computer Science and Engineering, Hindustan Institute of Technology and Science, Chennai, India.

The study is all about career guidance to the students on social networking sites.

Facebook, exploring the communities between the students within the group By Post, By tag, By share, By like, By unlike, frequency of messages were carried and tabulated by finding the strong ties and weak ties, For analyzing the ties if it is week we will give more suggestion to improve the guidance between the users, for the students we will give awareness of the studies and we understand the psychological traits of the students based on the aspects and we make the strong links between the student. In our study investigated the relationships in Social Media Network (SMN): Facebook (FB) and what were the possible mining techniques to explore and predict links in the network. Our proposed study interacts with the social network plays a significant role in learning knowledge engineering from students through cognitive collaboration with nodes. The SNA performed by the software application tool which predicts some important information related to a member's participation, interaction-structural analysis between them within the network. For the higher education perspective, the study explored in various latest technologies that can be used on campus so that Higher Education gets the facelift by going social.

II. RELATED WORKS

An optimal algorithm has used to predict the student's performance in academics and the framework was designed to help students to identify factors [1]. Nguyen et al., have proposed performance determination of the students, values were compared with logistic and linear regression predicted results were tabulated in their paper[2]. Kanokwan presented a paper for importance for career guidance opportunities for students [7] but his analyzed values are leads to a low retention rate. [24][23] Found the prediction of links in the complex and dynamic nature of links and yielded the good accuracy value. [8][9] were find accuracy value by used naïve Bayes classification.[20] correlation and clustering techniques were to used group the variables [11].[12] have proposed the emerging techniques of Link Prediction in the relationship between nodes by a supervised learning strategy. [13] Have analyzed the similarity between a pair of nodes. The total weight of the nodes and the set of constraints were assigned to one node. [19] Have proposed the structural patterns of the community behaviors in social networks were analyzed and predicted the accuracy of links.

[20] Have founded the correlation of edges and vertices, it helps to compare supervised and unsupervised links. [3] Have proposed and found the Community detection based similarity and degree of clustering in the same network. [11] Presented Sequential file access such as pre-determined order; Random files from a probability distribution, the file accessed using random walk prediction and Gaussian technique. These were supported for link prediction in social networks [18]. [17] Has presented the Community information uses the label propagation approach to detect different communities clustering in the social networks. The Bayesian theorem is used to detect the missing or broken link in the community networks [19]. Proposed the strength of the projection model in the network and proximity measures of bipartite networks [11] they used time-aware based on measuring the proximity of nodes. [17] proposed the content information while considering all other structural information, but it is depending on network structure and would not be appropriate for other networks. It is also having scalability issues because of their time complexity [5]. [20] proposed a hybrid method is a train the classifier by incorporating topology and contextual information. The weight measure approach uses topology information to predict the link between two nodes. It identifies the link exists between the two nodes using their profiles. The weighted score is quantified to predict the links. It considered the local structural information only and the user-generated contents are ignored which makes this method to use in the limited environment. The hybrid link prediction method integrates the content and structural information of the topology. It uses a topic inclusion degree to measure the relatedness of two users. The factorization matrix is used to build the newly constructed network which derives the relatedness of each node in the graph. This method is not considered the global structural information and it is not effective in the real large scale social networks. [22] have proposed predictions of binomial, replica replacement algorithms are used to common in all environments [16][15] it is the advantage of utilizing replicas in the link. [8] Have proposed Bayes classification with linear discriminant analysis. These techniques supported for analyzing the links.

III. CLUSTERING

By using social network analysis tools gives the statistical information. This analysis gives us the frequency of postings by each user. Many researchers are using traditional approaches to analyze the member's active participation likes questioning, text analysis methods which cater depth analysis of textual information shared among the students or other members of the group but neglect the structural aspect of the network. To analyze the degree of interactions between actively participating students, the study needs to investigate the network thoroughly by using SNA tools which predict the information about the structural properties, identify communities', to detect relevant clues for predicting active participation and finally the social relationship patterns.

IV. FEATURE EXTRACTION

Now the study will talk about commonly used web-enabled technologies i.e., SNS which is popular with the younger generation, especially higher education students and Academicians. These have much higher acceptance rates to these SNS platforms. The characteristic features of these SNS are (i) Sharing Content, (ii) To Communicate with peers, (iii) To develop Community relationships (iv) valuable knowledge resource.

Another feature of SNS is to have a social linking, This SNS builds friendships or social relationships not only between student-student connections but also among faculty-student or Higher Education Program e-student relationships. it was shown that many HEI's develop their E-learning environment to link actively their current or aspirant students, parents and the Higher Education Programs. This kind of modeling architecture of SNS gives us a feasible solution to the HEI's or persuaded educators to keep their Higher Education Program's updated and to have a positive bonding with their students.

Our study discussed in detail about the Facebook social networks The study found that these networks consist of actors (individuals, organizations, countries or any entity, etc., who are associated with each other by one or more edges or connection(s) (such as links, collaborations, citations, etc.,) The SNA is a process to examine and explore: (a)Parameters of the community (b) Links (c)To understand Node (d) Distortion Node (e) Density (f) Shortest path (g) In-degree (h) Out- degree and (i) to Identify cluster.

V. THE MATHEMATICAL FORMULATION OF CLUSTERING

The clustering formulation is to explore the strong connections between the links and edges and find the accuracy of link prediction in the clustered network. The following metrics were to calculate the aspects of career guidance to the Higher Education Students.

$$Z = \{Z1, Z2, \ldots\ldots\ldots Zn\} - n\ records\ of\ data$$

$$Z1 = \{Z11, Z12, \ldots\ldots., Z1n\}\ dimensional\ vector\ of\ data$$

$$Yj = Cluster\,(Xj) = \arg j \min \|Xi - \mu j\|$$

$Disortion = \sum_{j=1}^{k}(Xi - Ci)^2 = \sum_{j=1}^{k}(Xi - \mu j)^{\wedge}2$with in cluster.

VI. ACCURACY OF PREDICTION EXPERIMENTAL RESULTS

DATASET: User records are selected due to the calculation. This data gives 6000 nodes for the Post, 5540 for Likes and 3000 for Shared Count, The Links of the nodes in the Spatial Data around 30,000 to 55,000 links, Calculated values of Average Shortest Path and Global Clustering Co-Efficient were tabulated below.

Table 1: Career Guidance Facebook Dataset

	Post	Like	Share
Nodes	6,000	5,540	3000
Links	52000	45000	39000
GCC	0.026	0.0278	0.6564
ASP	3.00	3.068	163.76

6.1 Global Clustering Coefficient

It calculates the transitivity of the clustered network. The Co-Efficient range from 0 to 1.

$$GCC = \frac{3 * No\ of\ triangles\ in\ the\ cluster}{Number\ of\ connected\ vertices\ in\ the\ cluster}$$

6.2 Average Clustering Coefficient

Clustering is based on the Communities of a vertex in the network.

$$ACC = \frac{Number\ of\ connected\ vertices}{Number\ of\ triangles\ centered\ on\ a\ vertex}$$

6.3 Network Density

The ratio between real connection of existed edges and possible connection edges within a given network.

$$Density = \frac{Number\ of\ existed\ connected\ edges}{Number\ of\ Possible\ edges \quad // (v * v - 1)}$$

6.4 Average Shortest path

The paths are most likely to be between each pair of vertices. Its measures the clustering coefficient and its degree distribution.

$$\frac{1}{v * (v-1)} \sum_{i=j} d(i,j)$$

By these four methods calculate the accuracy of prediction for the career guidance aspects and it had been shows the correlated values below.

Table 2: Accuracy of Prediction values

Datasets	GCC	ACC	Network density	Gini Coeff
Scholarship	0.0254	0.1176	0.000677	4.672
Subject Ranking	0.058	0.3294	0.0119	2.254
Academic Reputation	0.019	0.1075	0.0075	3.236
Sports	0.003	0.2666	0.00079	4.186
Location	0.03	0.1946	0.002	2.581
Weather	0.017	0.2838	0.0029	3.277

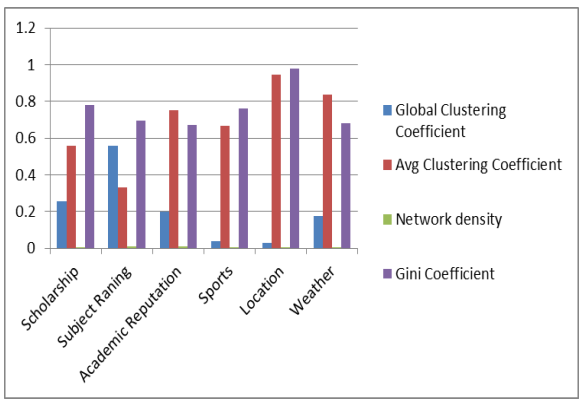

Fig.1 Prediction Accuracy Metrics

This graph depicts four different algorithms Global Clustering Coefficient (GCC), Average Clustering Coefficient (ACC), Network Density (ND) and Gini Coefficient (GC) used to analyze the university aspects for the students. GCC for the subject ranking was the highest source 0.5 of university aspects, In the next GCC for the scholarship 0.3 reaching half of the subject ranking, Other categories each represented about weather 0.15, Academic reputation as0.18 not as higher than the scholarship. Among the university aspects of GCC Subject Ranking is the main source for the students, this feature reaching half of the intake. ACC reached 0.9 for the location and 0.8 for the weather respectively. Academic reputation increased to over sports and scholarship and even slightly less than the weather. ND for the connected links subject ranking reached the strongest connectivity of all the nodes 0.1 remaining aspects slightly less than the subject ranking.

GC shows the preference of Location 1.0 and subject ranking 0.9. Overall for algorithms ACC and GC have a clear value of all aspects which varies depends on the features of data. However, the students give more preference for choosing the location.

Table 3: Predicted values of Algorithms

Dataset	K-Near Neighbour	Jacquard	Common Neighbour	Adami-c Adar	Cosine similarity	Leicht Holme Newman similarity
Scholarship	0.7	0.7	0.7	0.7	0.7	0.7
Subject Ranking	0.9	0.9	0.8	0.9	0.9	0.7
Acd Reptn	0.7	0.6	0.8	0.7	0.6	0.6
Sports	0.7	0.6	0.8	0.7	0.7	0.6
Location	0.8	0.8	0.9	0.8	0.8	0.8
Weather	0.9	0.8	0.9	0.9	0.8	0.8
Average	0.8	0.7	0.8	0.8	0.7	0.7
Variance	0.01	0.01	0.01	0.01	0.01	0.00

VII. THE AVERAGE CLUSTERING VALUES

The graph depicts about the mapped data between the students and university aspects, the chart shows the information of how much average frequency was the yield on the six different algorithms in the manner of Ease of Interpret (EoI) output, Calculation Time (CT) and Predictive Power (PP). According to the chart, all frequencies are in upward trends. In the Common neighborhood algorithm steadily increased by 0.88, nodes are likely to be similar and strongly connected in the cluster.

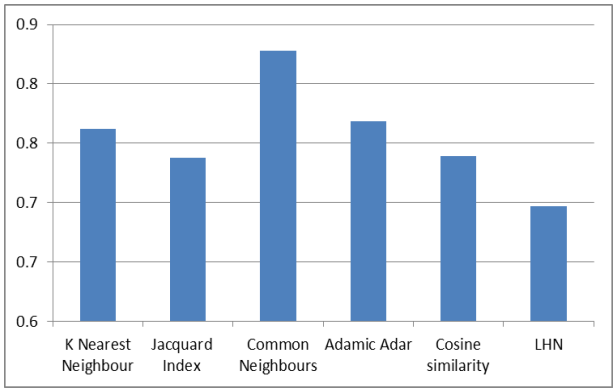

Fig.2 Average Clustering Values

Adamic Adar remained level at 0.80 and K nearest neighbor sharp decreased in 0.79 due to the calculation time and predictive power was lesser than the common neighbors. EoI was support to yield the gradual values but the overall frequency of average value declines at 0.1 level. In the next, the jacquard Index and Cosine similarity were less considerably used in the calculation time and predictive power. Although they are still less than the value of Common Neighbours. Cosine Similarity and Jacquard Index changed marginally in the value of 0.79 and 0.78. LHN yielded 0.7. Overall of these algorithms, common neighbors yield a good average frequency of all other average frequency values.

VIII. THE RELATIVE METHODS OF VARIANCE IN CLUSTER

This graph compared variance between the algorithms it is clear from the data in the aspects of Relative Performance, Distance and Computation Cost. Overall frequency ranges were increased in the level of 0.01 remained standability of all algorithms but the changes of fluctuation come into the common neighbors and LHN, the relative performance and distance of nodes weakly supported in the algorithms so it yielded 0.00005 and 0.0014 respectively. The neighbors are outlier in turning the algorithm of the computation cost is comparable to the cosine similarity 0.0012. It is interesting to note that K-Nearest Neighbour to increase the values of 0.0167 the relative performance and distance give more supported to find the neighboring nodes and edges between the nodes and shortest path of the nodes has easily found in the KNN algorithm. the computation cost where 0.8 were higher than the remained algorithms.

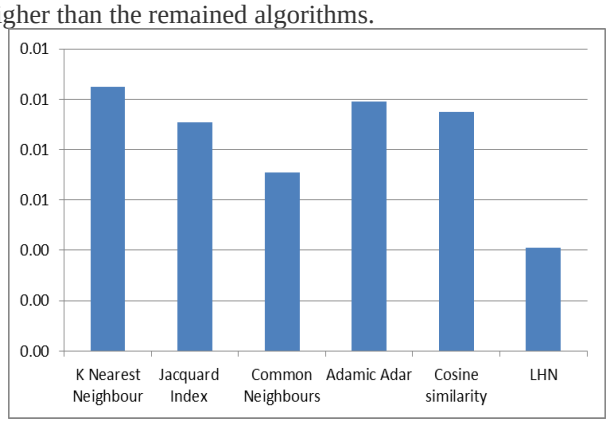

Fig.3 Relative methods of variance in cluster

In the conspicuous view jacquard index and Adamic Adar where 0.0017 and 0.0019 were comparatively supported to yield the values, but the relative performance of the jacquard index has lesser than the Adamic Adar, the computational cost value is better than the cosine similarity values. Overall the performance of Variance between the algorithm KNN algorithm sequentially supported to find the nearest neighbor and the distance of the nodes and the computational cost also yield good accuracy better than the remained methods.

IX. CONCLUSION

In this paper, we did the parallel relationship between the four prediction methods and different network metrics used to analyses the values of the spatial dataset. The result of the prediction accuracy shows that the best technique among the evaluated ones is K-Nearest Neighborhood Algorithm. It is the steady method for the tried ones, Common Neighborhood is the second technique, also gives good prediction accuracy. We extend this research for finding the relationship between too many other communities in the cluster which might be good for finding the strong ties relationship and more general relations.

REFERENCES

1. Kanokwan Kongsakun, Chun Che Fung, Tuchtawan chanakul, "Developing An Intelligent Recommendation System For A Private University In Thailand", Issues in Information Systems, Volume XI, No. 1, 2010
2. Nguyen Thai-Nghe , Lucas Drumond , Artus KrohnGrimberghe , Lars Schmidt-Thieme, "recommender System for Predicting Student Performance", 1877-0509 @ 2010 Published by Elsevier B.V. doi:10.1016/j.procs.2010.08.006 Open access under CC BY-NC-ND license. Open access under CC BY-NC-ND license
3. Bolen Chen, Ling Chen "A link prediction algorithm for colony optimization"Springerlink 2014.
4. Bryan Perozzi, Rami Al-Rfou,and Steven Skiena. Deepwalk:Online learning of social representations. InKDD.ACM, 701–710. 2014.
5. Dang, Bo Wei. Ya. "A Cluster Growing Dimension of Social Network:For closeness Centrality." ELSIEVER. PHYSICA.A, 2019.
6. JianTang,MengQu,MingzheWang,MingZhang,JunYan,and Qiaozhu Mei. Line:Large-scale informational work embedding.In WWW.ACM, 1067–1077. 2015.
7. T. Sudalai Muthu and R. Masillamani, "A novel protocol for secure data storage in Data Grid," *2010 First International Conference On Parallel, Distributed and Grid Computing (PDGC 2010)*, Solan, 2010, pp. 184-189
8. Vijayalakshmi Karthiga S., Sudalai Muthu T., Roberts Masillamani M. (2014) "Early Detection of Diabetic Retinopathy by CRA (Cluster Rejection Approach)". In: Sathiakumar S., Awasthi L., Masillamani M., Sridhar S. (eds) *Proceedings of International Conference on Internet Computing and Information Communications*. Advances in Intelligent Systems and Computing, vol 216. Springer, New Delhi
9. Rani, V.K. & Sudalai, M.T.. (2015). "An innovative design of intelligent cradle for infants". International Journal of Applied Engineering Research, vol. 10, issue.8, 2015, 20867-20873.
10. T. S. Muthu and K. R. Kumar, "A Log-based predictive approach for replica replacement in data grid," *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore*, 2017, pp 1-6.
11. T. SudalaiMuthu and K. RameshKumar, "Hybrid predictive approach for replica replacement in data grid," *2017 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore*, 2017.
12. T. S. Muthu, R. Vadivel, A. Ramesh and G. Vasanth, "A novel protocol for secure data storage in Data Grid environment," *Trendz in Information Sciences & Computing(TISC2010)*, Chennai, 2010, pp. 125-130
13. Pandiaraj, S. and Sudalai Muthu, T., "Prioritization of replica for replica replacement in data grid", *International Journal of Recent Technology and Engineering*, Volume 7, Issue 5, February 2019, Pages 245-248.

14. Keith Henderson, Brian Gallagher, Tina Eliassi-Rad,: structural role extraction&mining in large graphs. In KDD. ACM, 1231–1239I.
15. Liben-Nowell. D and J. Kleinberg, "The link-prediction problem for social networks," J. Am. Soc. Inf. Sci. Technol., vol. 58, no. 7, pp. 1019–1031, 2007.
16. M. A. Hasan, V. Chaoji, S. Salem, and M. Zaki. "Link Prediction Using Supervised Learning." Link Analysis Counter Tourism and Security. SDM, 2018.
17. Mark Heimannand Danai Koutra. "On Generalizing Neural Node Embedding Methods to Multi-Network Problems" InKDD MLG Workshop 2017.
18. Muthu.T. S and Kumar. K "Hybrid predictive approach for replica replacement in data grid", International Conference on Advanced Computing and Communication Systems, Article number 8014706, August 2017.
19. Pandiaraj. S and T. SudalaiMuthu " Prioritization of replica for replica replacement in data grid", International Journal of Recent Technology and Engineering Volume 7, Issue 5 2019.
20. Ranjana, P, Lakshmi Sridevi, S. Sudalai muthu. T, Vikram Gnanraj, "Machine Learning Algorithm in Two wheelers fuel Prediction" International Conference on Innovations in Information and Communication Technology, April 2019
21. Serpil Aslan, Buket Kaya, Mehmtan. "Predicting Potential Links By Using Strngthened Projections in Evolving BiParite Networks ." PHYSICA A. Stastical mechanics and its Application, 2019.
22. Rohini.A Leelamabika. K. V Queen Marry Vidya, "Bayes Classification for fingerprint retrieval 2013.
23. Rohini.A SudalaiMuthu T., "A Weight based Scheme for Improving the Accuracy of Relationship in Social Network", International Journal of Innovative Technology and Exploring Engineering., vol. 8,Issue 11,2019.
24. Xiao Huang, JundongLi, andXiaHu. Label informed attributed network embedding. InWSDM.ACM, 731–739. 2017.