Assessing Research Collaboration in Database Systems and Computer Networks by Analysis of Coauthorship Network

Adeel [1]
Department of IT
The University of Haripur, Haripur, KPK,
Pakistan
aahmedqau@gmail.com

Tanveer Ahmed[2]
Department of Computer Science
The University of Lahore, Gujrat,
Pakistan
tanveir.ahmed@hotmail.com

Abstract— **Community detection is a fundamental problem in social networks. These networks detect communities based on link analysis and strong connection strengths, but cannot reflect Author's from different research areas. To address the problem of community detection, we have done a study for "Analyzing patterns of collaboration in co-authorship network using Modularity and Centrality Measures". This analysis study uses combine features of Modularity with centrality measure to effectively detect community of different author's having different research collaboration with different interests in domain of Computer Networks and Database Systems. Experiment of Dataset shown that this approach is better detect best authors from specific domain having high collaboration with other coauthors and presents information to the researcher's that have relative interest in relative author's community.**

Keywords—social network analysis; centrality measures; community detection; complex network; undirected graph

I. INTRODUCTION

The study of social networks analysis in the past reveals patterns of collaborations that helps better understanding and decision making [1]. Analysis of this network reveals patterns of publications by identifying most prominent researcher's communities. Research publications brings citation for authors and it also establishes reputation of an authors in coauthorship network. The reputation of author gets better sponsored research as well as continuation of efforts in research. Researcher's scientific publication shows interest in particular domain or research field. In this analysis study author and co-authors communities have detected with different collaboration patterns. Community is the structural component of the networks that describe relations and interaction between nodes. Communities can be represented with different names such as groups, modules, clusters etc. In vertex sense, all communities are considered as group of vertices which are similar to each others [2,3,4,5,6,8,9].

Graphical visualization of coauthorship network of database systems is shown in Fig. 1 and computer networks is shown in Fig. 2.

The organization of this papers is as follow. Section 2 describe about existing work, section 3 describe about dataset and experimental results and in section 4, we discuss some existing methodology and conclusion.

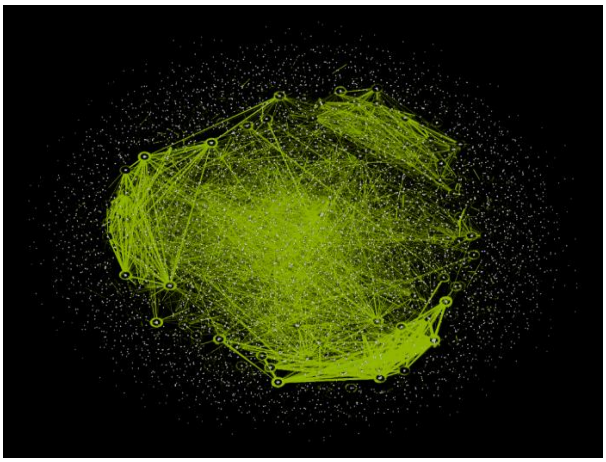

Figure 1 Co-authorship network of Database Systems

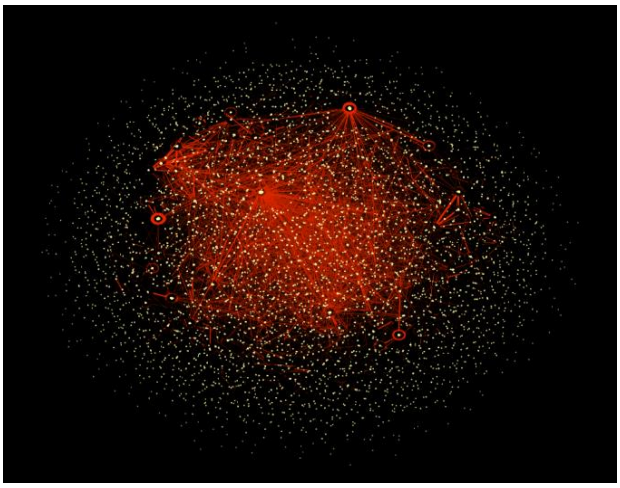

Figure 2 Co-authorship network of Computer Networks

II. Related Work

Methods and algorithms have been developed over time use different approaches, most of them uses nodes to identify how these are linked with other nodes. Nodes that are mostly linked with other nodes are identified through centrality measures based on some similarity [10].

In computer networks, centrality might be used to identify a router through which data can pass through easily via shortest path and also this reveal information that this router can be most vulnerable to viruses and threats. Most important similarity measures are:

Degree Centrality.

Degree centrality represent an actor or node that have most number of contact such as links or relations with other actors in the networks. High degree centrality of an actor show that the actor is most influential in the network [11,12].

Closeness Centrality.

Closeness Centrality works on the geodesic distance that a node has with other nodes in the network. Closeness of a nodes states how long information will take to reach other node in the network. A node that is close to other nodes in the network can communicate with other nodes in the network [13].

Betweenness Centrlaity.

Betweenness Centrality of an actor or node describes that a node lies between two other nodes while communicating with each other. Number of times a node visited between two other nodes shows that visiting node has high betweenness centrality value. The node with betweenness might control the flow of information between other nodes because it is considered to be more central for communication for other nodes. Same betweenness of different nodes might results in the probability of being used at an equality level.

PageRank.

PageRank used by Google and it is proposed by Page and Berin. It is used to rank webpages in search engine results. PageRank rank any webpage on the basis of factors such as number and quality of webpage by assigning weight to webpages that determines how much importance a website has.

In this paper, we have Analyzed patterns of Author's and Co-authors in Co-authorship network based on Modularity and Centrality Measures. This Analysis study reveals Author's community that have best publications. In this proposed work, we have used Modularity and Centrality Measures for detecting best authors and co-authors communities.

There is a growing interest of Data Mining and Text Mining research in the development of community detection methods for different Social Networks using different Optimization technique as it is a very important contribution in finding best community structures. The main aim is to detect communities of different social network sites that have different structures and provides better information to society.

III. Dataset And Experimental Results

For applying four most widely used classic measures (closeness centrality, degree centrality, betweenness centrality and pageRank) to co-authorship network, Microsoft Academia Research [14] dataset has selected in this research. Microsoft Academia Research is an experimental service developed by Microsoft that explore how authors, students, scholar and researchers find contents. It also shows relationships among subjects, author and contents. The description about Database Network and Computer Network are shown in table 1. There are 26249 authors in Database network, in which these authors have written 62975 papers, network has modularity 0.904, network diameter is 26, network density is 0.001 and connected components in database network are 5964. Similarly, there are 19241 authors in computer networks in which these authors have written 43831 papers, network has modularity 0.895, network diameter of computer networks network is 27, network density is 0.001 and connected components in database network are 5579. Network diameter in graphs shows longest path in the graph. In Database the longest path is 26 and in Computer network longest path is 27. In undirected graphs connected component is subgraph in whole graph. 5964 vertices in Database graph are connected with each other while in Computer networks these connected vertices are 5579.

Table 1: Coauthorship Network Statistics

Statistics	Database Network	Computer Network
Number of papers	62975	43831
Number of authors	26249	19241
Modularity	0.904	0.895
Network Diameter	26	27
Network Density	0.001	0.001
Connected Components	5964	5579

A) Apply degree centrality

Figure 3 shows top 30 authors graph of nodes of database and computer network based on degree centrality. To understand which authors has highest degree ranking is assigned on the basis or degree value. Same degree centrality shows same rank in the network. Few authors in top 30 computer network network authors have same degree value such as marco tacca an jun wang has 114-degree value and ranked 19, William j mitsh, c staver, john e bowers and david p hill has 113-degree value and ranked 20. In database network eds has highest degree value 822 and has ranked 1, e somarriba has 342-degree value and has ranked 2. On average in top 30 authors, c a Harvey and l hilje has 177 degree and 13 rank, At the bottom a rodriguez Navarro, s pascual Tovar, j a Alonso velasco, v r Gonzalez fernandex, s blanco arenal, I gutierrer montes, s blanco suarez and r escribano romo has 154 degree and 21 rank.

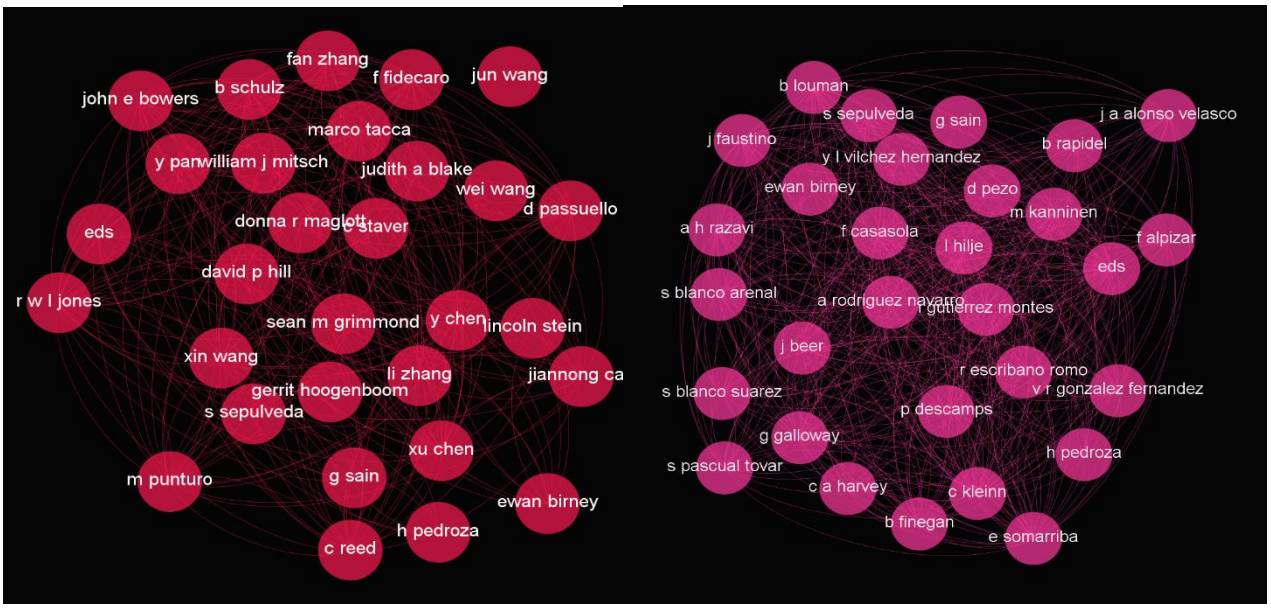

Figure 3 Degree centrality of Computer Networks and database

(B) Apply closeness centrality

In computer network, wei wang has highest closeness which is 0.1789 and also it has high rank in computer network which is 1, jie zhang has 0.1709 closeness value and has ranked 2. In the middle yongli zhao has 0.1646 closeness value and his rank is 13, li zhang has 0.1628 closeness value and has ranked 21. At bottom jie li, weisheng hu has same closeness value 0.1623 and rank 25, jun yang, xin wang and yu liu has 0.1619 closeness value and 26 rank. Few authors have similar closeness value; their shortest path is different same to other coauthors. Similarly, in database network, Wei wang has 0.1754 closeness and 1st rank, jun wang has 0.1690 closeness and 2 rank, jie zhang has 0.1680 closeness and 3rd rank while jian wang has 0.1666 closeness value and 4th rank. In middle y pan has 0.1637 closeness and 11th rank, fan zhang has 0.1630 closeness and 15 rank. Author hui li and dong wang has similar closeness and rank (0.1622 and 17, at bottom eds has 0.1602 and ranked 25. Figure shows top 4 of undirected graph for closeness centrality.

C) Apply betweenness centrality

Betweenness centrality of a node represents how diversely published paper with other coauthors. In the last of top 30 authors, ying li has 0.00726 betweenness and 29 rank while jing li has 0.00721 betweenness and 30 rank. No author has similar betweenness value. This represent that there is variation in betweenness value in computer networks network. In database network, eds has 0.03993 betweenness and 1st rank, wei wang has 0.03548 betweenness and 2nd rank. In middle bo liu has 0.00900 betweenness and 15 rank, yun li has 0.00875 betweenness and 17 rank, fan zhang has 0.00839 betweenness and 20 rank. At bottom, elisa bertino has 0.00684 betweenness, young min kim has 0.00683 betweenness and ranked 29 and 30 respectively. Figure 5 shows graph of top 30 authors for betweenness degree.

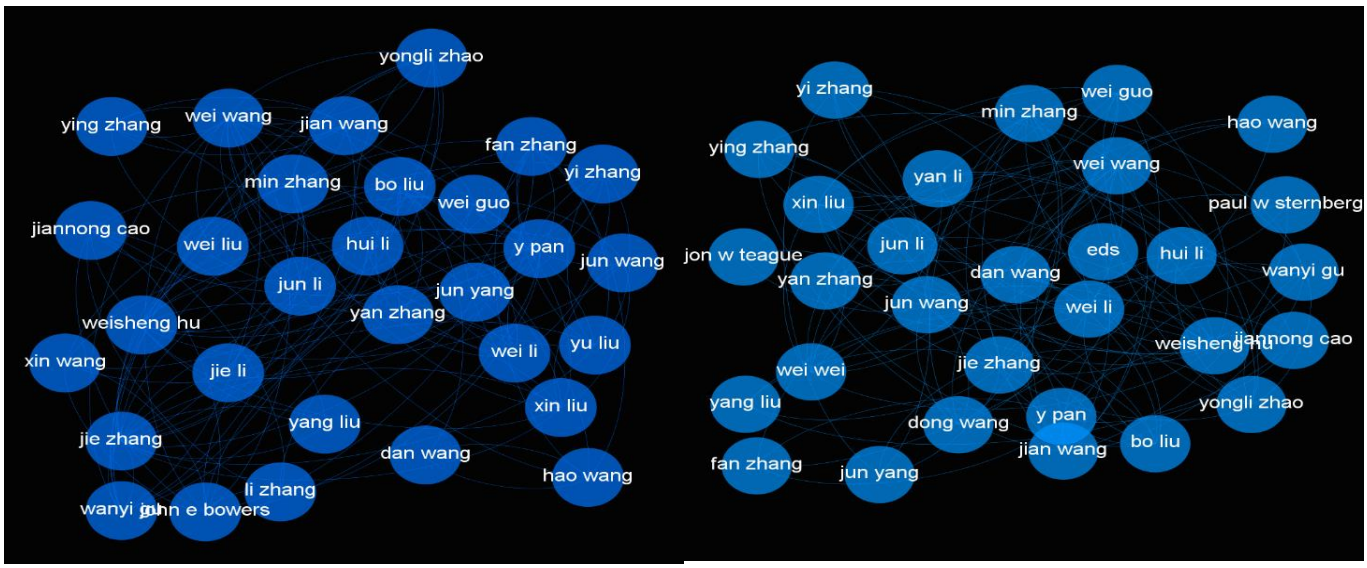

Figure 4 . Closeness centrality of Computer Networks and Database

D) Apply PageRank measure in coauthorship network

PageRank represents how productively author has published research papers with other coauthors. In computer network, eds has highest PageRank 0.00325 and also it has high rank in computer network which is 1st, rodrigo munguia has PageRank 0.00120 and ranked 2nd.

In database network, wei wang has ranked 1st, because he has 0.00079 PageRank value and eds has 0.00066 PageRank value and ranked 2nd. In the middle few author has same PageRank, wei liu, yuefeng ji and lei wang has PageRank 0.00034 and their ranking is 8, jun li, ioanniz tomlos, wanyi gu and hui li has PageRank 0.00033 and ranked 9. At bottom biswanash Mukherjee, yang liu and jian zhang has 00027 PageRank and ranked 15. Figure 6 shows top 30 authors of undirected graph.

IV. Conclusions

In this paper, we have performed an analysis study on coauthorship networks of Database Systems and Computer Networks using centrality measures. We ranked the authors according to domain and highlighted the most important authors in the field. In future, we will perform analysis on other domains using machine learning techniques.

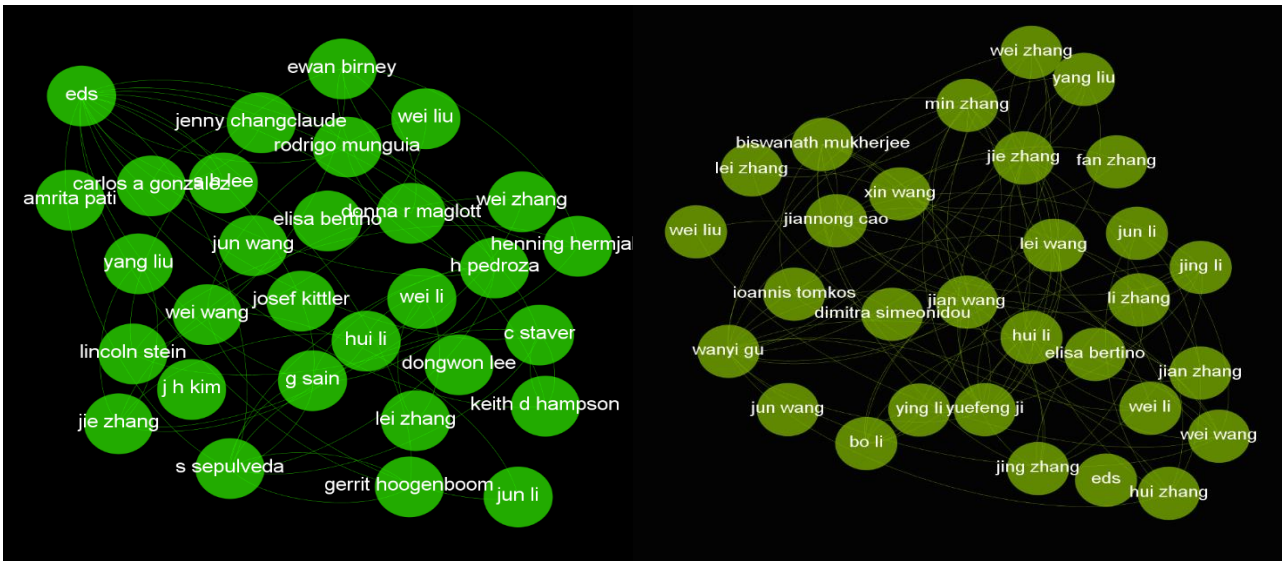

Figure 5 Betweenness centrality of Computer Networks and Database

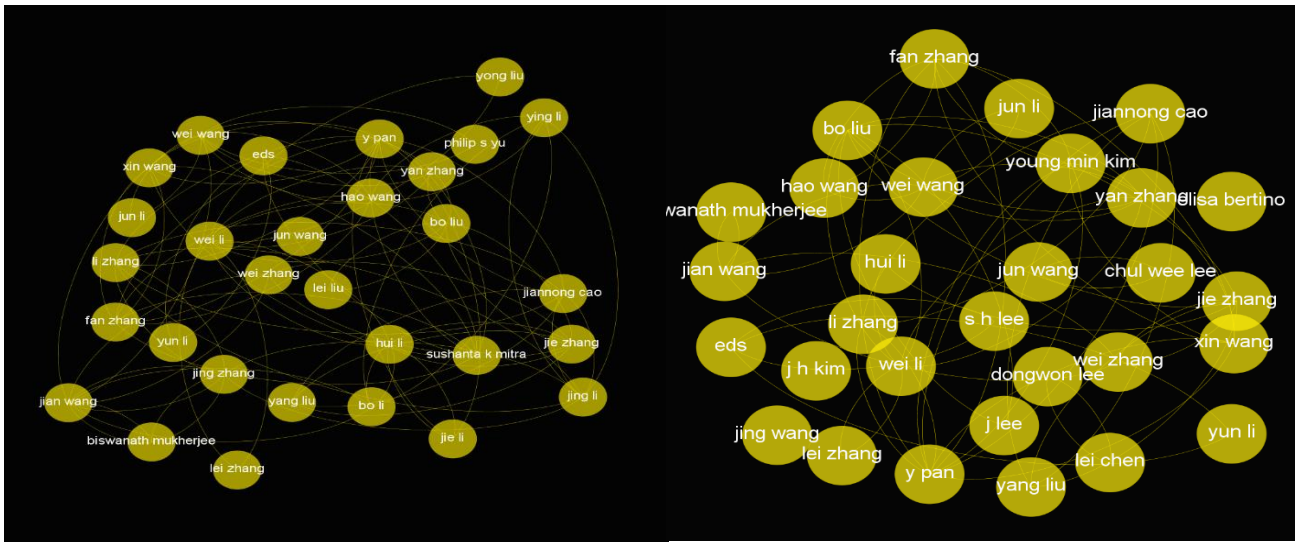

Figure 6 PageRank of Database and Computer Networks

REFERENCES

[1] K. Sutaria, D. Joshi, C,K.Bhensdadia and K.Khalpada, "An Adaptive Approximation Algorithm for Community Detection in Social Network", Computational Intelligence & Communication Technology (CICT), IEEE International Conference on, 2015.

[2] X. Xhang, H. You, W. Zhu, S. Qiao, J. Li, L. Alberto, Z. Zhang and X. Fan."Overlapping community identification approach in online social networks". Physica A: Statistical Mechanics and its Applications, 421: 233-248, 2015.

[3] M. E. J. Newman. "Modularity and community structure in networks", Proceedings of the National Academy of Sciences, USA, 103(23): 8577-8582, 2006.

[4] Z. Zhao, S. Feng, Q. Wang, J. Huang, G. Wlliams and J. Fan, "Topic oriented community detection through social objects and link analysis in social networks", Knowledge-Based Systems, 26: 164-173, 2012.

[5] F. A. Cheong and J. Brian, "A Social Network Analysis of the Co-Authorship Network of the Pacific Asia Conference on Information Systems From 1993 To 2008", The Pacific Asia Conference on Information Systems Pacis, 2009.

[6] Umadevi, D. V., "Case Study – Centrality Measure Analysis On Co-Authorship Network", Journal of Global Research in Computer Science 4(1), 2013.

[7] E. Y. Li, C. H. Liao and H. R. Yenn"Co-authorship networks and research impact: A social capital perspective". Research Policy 42(9): 1515-1530,2013.

[8] S. Suhas and S. M. S Thorat. "A Survey on Community Detection", International Journal of Science and Research (IJSR) 4(1), pp:670-673,2015

[9] J. Zhang and P. S. Yu, "Community Detection for Emerging Networks". Proceedings of the 2015 SIAM International Conference on Data Mining: 127-135, 2015.

[10] X. Liu, J. Bollen,M.L.Nelson and H.Ven de Sompel, "Co-authorship networks in the digital library research community", Journal of Information Processing & Management 41(6): 1462-1480, ,2005.

[11] V. Dequiedt, "Local and Consistent Centrality Measures in Networks". Social Science Electronic Publishing, Inc., Social Science Research Network, 2015.

[12] J. Sun, and J. Tang C. Aggarwal, "A Survey of Models and Algorithms for Social Influence Analysis". Social Network Data Analytics. Boston, MA, Springer US: 177-214, 2011.

[13] K. Nidhi and K.M.P. Chitroda, "Improve Page Rank Algorithm using Normalized Technique", International Journal of Computer Science and Engineering (IJCSE) 3(4), 2014.

[14] A.W.Harzing, "Microsoft Academic (Search): a phoneix arsisen from the ashes", Journal of Scientometrics, Vol 108, 3, pp:1637-1647, 2016.