

IDENTIFICATION OF LEADING RESEARCH CONTRIBUTORS WITH NOVEL PERFORMANCE METRICS USING ACADEMIC SOCIAL NETWORK

G.Ayyappan¹, Dr.C.Nalini², Dr.A.Kumaravel³

Research Scholar, Department of Computer Science, Bharath University¹

Professor, Department of Computer Science, Bharath University²

Professor & Head, School of Computing, Bharath University³

ayyappangmca@gmail.com¹, drnalinchidambaram@gmail.com², drkumaravel@gmail.com³

Abstract

Social network is a structure of human relations and association. It is made up of a organized social actors in a network form. Information has varied number of forms and various purposes for communication. Journals serve as major source of primary information. The Researchers tend to publish more and more research output in journals. This paper focuses on the novel techniques for identifying the leading research contribution in the clusters of research social networks.

Key Terms: Data Mining, Academic Social Network, Weighted Frequency, Individual Frequency.

1. INTRODUCTION

In this section we focus on a different aspect of identifying the leading contributor as there is a relative lack of theory-based leadership studies in an R&D context (Scott & Bruce, 1998). More recent research has moved beyond simple examinations of leaders' traits and behaviors and we in particular consider the exchanges with the members, frequency of publications, participation at conference, etc.

Social network is a structure of human relations and association. It is made up of a organized social actors in a network form. Information has varied number of forms and various purposes for communication. Journals serve as major source of primary information; researchers tend to publish more and more research output in journals. Jie Tang et al.,[1] Extraction and mining of academic social networks aims at providing comprehensive services in the scientific research field. An academic social network, people are not only interested in searching for different types of information (such as authors, conferences, and papers), but are also interested in finding semantics-based information (such as structured researcher profiles). During the last few decades observed a dramatic development in co-production of research results. Almost every paper in the research field is co-authored.

The Section one introduces the topic of research and explains need for the study. The section two deals with the review of Literature pertaining to the research problem. The section three, deals with the materials pertaining to the research problem. The section four, deals with analysis of empirical methods and identifying leading research contributors using novel metrics. Finally section five, consist of the conclusion.

II. RELATED WORKS

In this section presents "Social Network Analysis of Author Collaboration" using DBLP, ACM Digital Library, CiteSeer, and SCI in Aminer[8,9]. Nasir Ahmad Aziz et al. [2] introduced a simple harmonic weighing algorithm for correcting citations and citation-based metrics such as the h-index for co-authorships. This weighing algorithm can account for both the number of co-authors and the sequence of authors on a paper. The concepts and methods introduced here may thus provide a more fair impression of a scientist's autonomous academic performance.

Jie Tang et al.[3] proposed to simultaneously model topical aspects of authors, research articles, authors and publication locations. Search services such as expertise search and people association search have been provided based on the modeling results. Jie Tang et al.[4] proposed the Cross-domain collaborations exhibit very different patterns compared to traditional collaborations in the same domain. Jie Tang et al.[5] proposed and developed The extracted profiles have been applied to expert finding, an important application on the Web. Jie Tang et al.[6] defined a two-step parameter estimation algorithm. Ash Mohammad Abbas [7] proposed and developed positionally weighted and equally weighted of authors for the research community.

III TERMS AND METHODS

Social network analysis

A “social” network is defined as a group of collaborating (or competing) entities that have some type of relationship and interact within a shared environment often referred to as a community.

Author collaboration

Research collaboration or Author collaboration can be defined as the working together of researchers to achieve the common goal of producing new research knowledge.

The dataset named as topic_paper_author in the academic social network data from AMiner [8,9]. The dataset is collected for the purpose of cross domain recommendation. The attributes contain Data Mining, Medical Informatics, Theory, Visualization and Database areas.

In this research work used Weka 3.6.9 [11], open source software for Text Mining process, MeSH [12] for identification of domain and SAS University Edition [13,14] which is getting permission to access SAS Studio from SAS Institute for Mining of the research community. Based on this novel metrics, the “Top most Influential researcher” of research community has been identified.

IV EXPERIMENTS AND RESULTS

The data has downloaded from https://aminer.org/topic_paper_author. The dataset is collected for the purpose of cross domain recommendation. Data Mining, Medical informatics, Theory, Visualization, Database.

There are 5 attributes in this dataset Conference Name, Title, Authors, Year and Abstract. We convert all abstracts in to topic. We made topic model processes, we converted all abstracts in to topic for using **MeSH (Medical Subject Headings, Topic Extraction Tool) of U.S National Library of Medicine**. We fix the class name is topic.

We describe the major tasks in social network analysis data sets used for studying this problem. We find authors belong to which topic in the dataset. We made topic model processes, we converted all abstracts in to topic for using **MeSH (Medical Subject Headings, Topic Extraction Tool) of U.S National Library of Medicine**. The Collections of nodes are called each and every person who involved in the social network. The Topic distribution like as Algorithm, data mining, database, Artificial Intelligence, Clinical, Medical Imaging, Image Processing, Biomedical Informatics, Image Processing, Telemedicine.

Only take authors data in Topic_Paper_Author dataset. It has splited in delimited in excel sheet it has split in 1 to 32 authors in each record. Investigator investigated leading individual research contributors and their weighted research contributions using Statistical Analysis System University Edition (SAS UE) for mining process. After applying of SAS UE, the resultant authors list consists of 52255 authors in this dataset. Goal of this framework to find Individual Frequency (IF) and Weighted Frequency (WF) of authors in this dataset. Finally, this frame work is applying the Cartesian product or cross product concepts in Topic-Paper-Author dataset.

Formula to compute novel approach in citation to find most and effective contributor in research.

Weighted Frequency (WF)

$$\rho_i = \sum_{i=1}^n (1 / i^{th} \text{ position}) * i^{th} \text{ position of Author individual frequency}$$

Where i = Author positions in each conferences

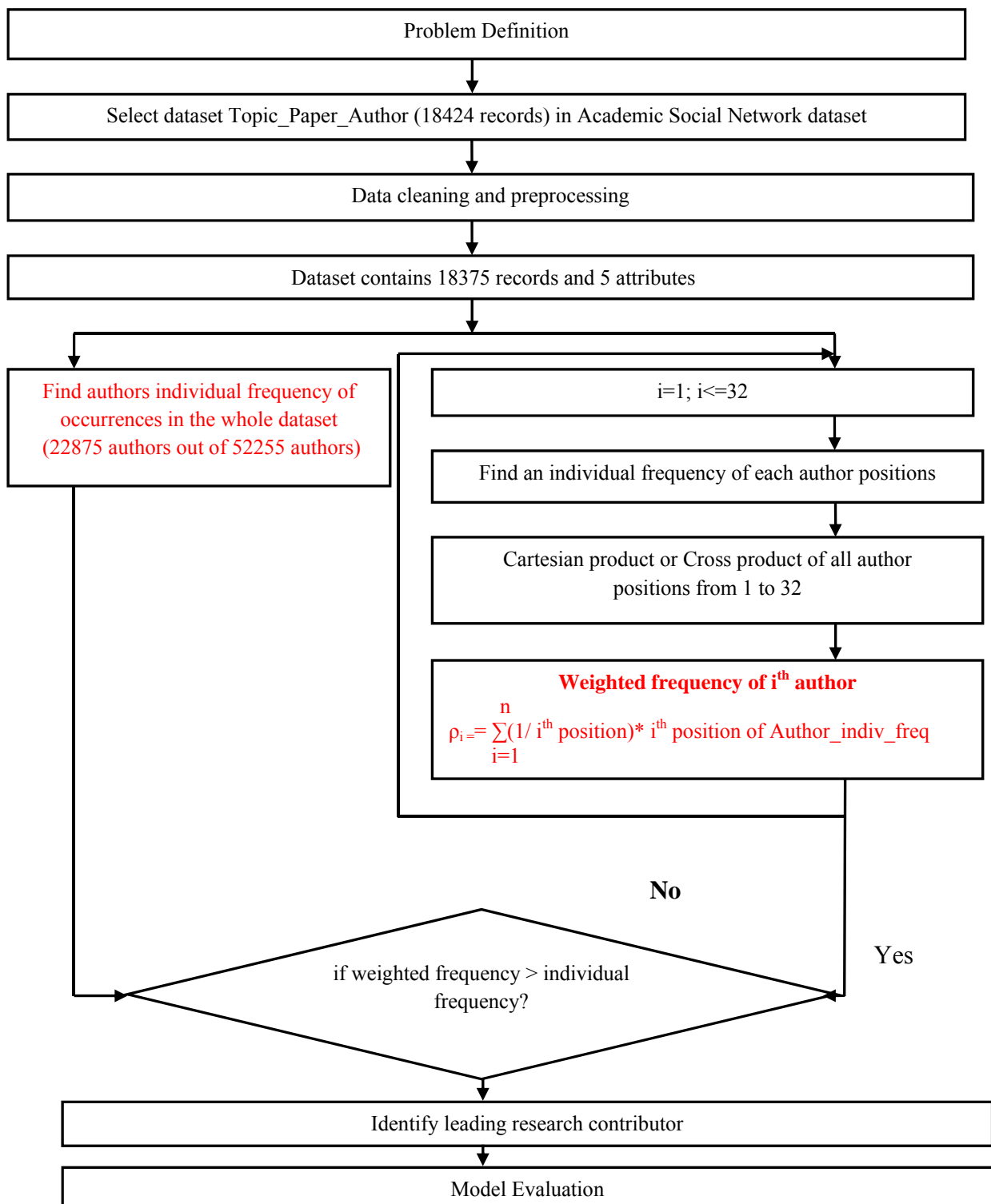


Figure 1 Proposed frame works to identify leading research contributors

In this study clearly emphasize on the leading research contributors. Surajit Chadhuri is having highest weighted frequency 57.58 but he is having 72 individual frequencies. Philip S.Yu is having 100 IF and WF is only 37.6. This below table representation of top twenty research computed based on propose system like as Weighted Frequency method. It clearly demonstrates their individual frequency also.

S.No	Author	Individual Frequency(IF)	Weighted frequency(WF)
1	Surajit Chaudhuri	72	57.58
2	Rakesh Agrawal	70	52.33
3	Jiawei Han	93	49.45
4	Noga Alon	47	46.5
5	Richard T. Snodgrass	60%	45.16
6	H. V. Jagadish	64	42.68
7	Michael J. Franklin	62	38.11
8	Baruch Awerbuch	40	38
9	Philip S. Yu	100	37.61
10	Hector Garcia-Molina	78	36.81
11	Michael J. Carey	51	34.82
12	Michael Stonebraker	55	34.56
13	Christos Faloutsos	68	33.53
14	Jon M. Kleinberg	46	32.83
15	Piotr Indyk	48	31.95
16	Charu C. Aggarwal	33	31
17	Marianne Winslett	33	31
18	Pankaj K. Agarwal	33	30.83
19	Sudipto Guha	39	30.36
20	Kenneth A. Ross	41	30.29

Table 1 Top 20 researchers based on proposed system WF and IF

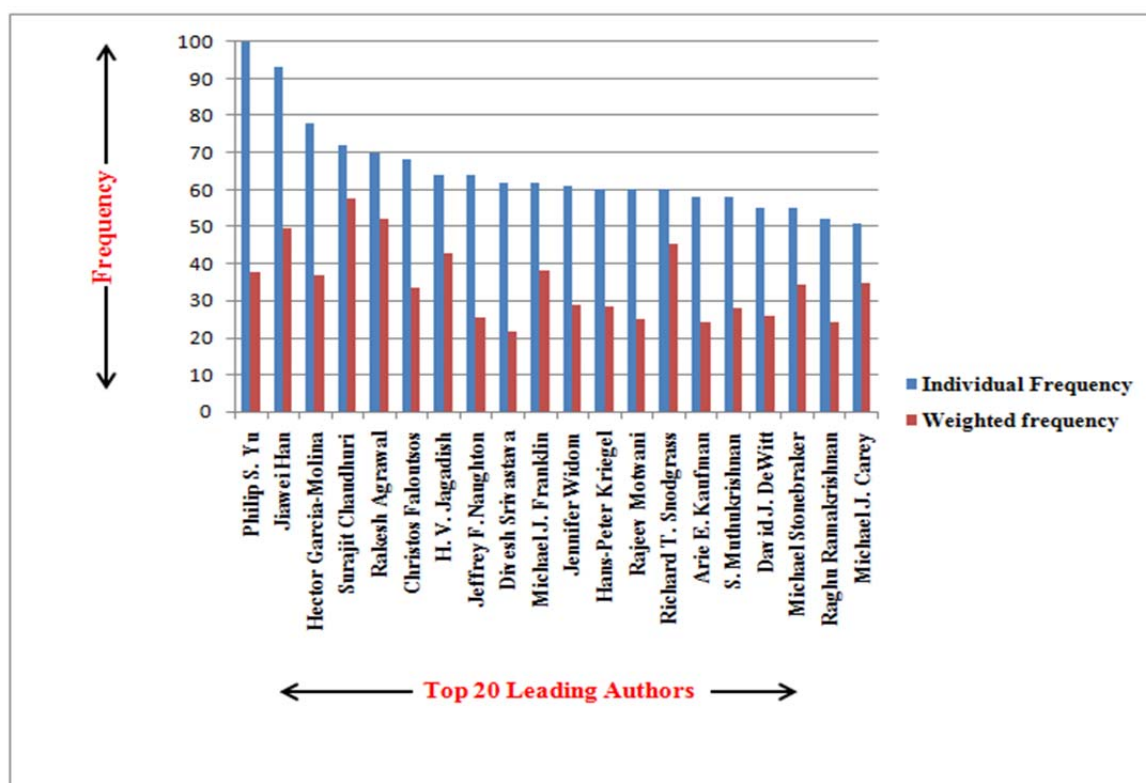


Figure 2 Top 20 individual frequency vs weighted frequency

Using the above formula Philip S.Yu is 100 research papers and the highest individual contributor but he is having weighted contribution is only 37.611. We take top 20 individual contributors Vs weighted frequency in this diagram.

V CONCLUSION

Based on this experiment easy identify who is the leading research contributor of the research society. Compute two methods one is identify each author's distribution of the dataset and other one is identify the weighted of the research authors in research area. Finally compare both weighted and individual authors distribution. The proposed framework for identifying leading researcher in based on the weighted distribution of authors belongs to authors position from namely first, second...and thirty second positions.

REFERENCES

- [1] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su, "Extraction and Mining of an Academic Social Network," *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discov. data Min.*, pp. 1193–1194, 2008.
- [2] T. P. Clement, "Authorship matrix: a rational approach to quantify individual contributions and responsibilities in multi-author scientific articles," *Sci. Eng. Ethics*, vol. 20, no. 2, pp. 345–361, 2014.
- [3] N. A. Aziz and M. P. Rozing, "Profit (p)-Index : The Degree to Which Authors Profit from," vol. 8, no. 4, pp. 1–8, 2013.
- [4] J. Tang, S. Wu, J. Sun, and H. Su, "Cross-domain collaboration recommendation," *Kdd*, p. 1285, 2012.
- [5] J. Tang, L. Yao, D. Zhang, and J. Zhang, "A Combination Approach to Web User Profiling," *ACM Trans. Knowl. Discov. Data*, vol. 5, no. 1, pp. 1–38, 2010.
- [6] J. Tang, A. C. M. Fong, B. Wang, and J. Zhang, "A unified probabilistic framework for name disambiguation in digital library," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 6, pp. 975–987, 2012.
- [7] A. M. Abbas, "Weighted indices for evaluating the quality of research with multiple authorship," *Scientometrics*, vol. 88, no. 1, pp. 107–131, 2011.
- [8] <https://aminer.org/data#Academic%20Social%20Network>
- [9] <https://aminer.org/data#Topic-paper-author>
- [10] <https://aminer.org/data>
- [11] <http://www.cs.waikato.ac.nz/ml/weka/>
- [12] <https://meshb.nlm.nih.gov/MeSHonDemand>
- [13] https://www.sas.com/en_us/software/university-edition/download-software.html
- [14] <https://support.sas.com/en/support-home.html>