Eigenvector centrality and its application in research professionals' relationship network

2 authors, including:

Anand Bihari
VIT University
24 PUBLICATIONS   **46** CITATIONS

Some of the authors of this publication are also working on these related projects:

Researcher's Scientific Assessment View project

Protein Classification View project

Eigenvector centrality and its Application in Research Professionals' Relationship Network

Anand Bihari
Department of Computer Science & Engineering
Silicon Institute of Technology, Bhubaneswar,
Odisha, India.
Email: csanandk@gmail.com

Manoj Kumar Pandia
Department of MCA
Silicon Institute of Technology, Bhubaneswar,
Odisha, India.
Email: manoj.pandia@gmail.com

Abstract—**The centrality of vertices has been the key issue in social network analysis. Many centrality measures have been presented, such as degree, closeness, between's and eigenvector centrality. But eigenvector centrality is more suited than other centrality measures for finding prominent or key author in research professionals' relationship network. In this paper, we discuss eigenvector centrality and its application based on Network x. In eigenvector centrality first set every node a starting amount of influence then performs power iteration method. In network x the starting amount of influence of each node is 1/len(G). Therefore, we modify the eigenvector centrality algorithm and set the starting amount of influence of each node is the degree centrality of that node because eigenvector centrality is the extension of degree centrality and also implements the eigenvector centrality in weighted network.**

Keywords—eigenvector centrality; social network analysis; degree centrality

I. INTRODUCTION

A research professionals' relationship network is a set of research professionals which has connections in pair to represent their relationship. Two researchers are considered in a relationship, if they have published articles in journal, conference and publish & edit books together. In such type of network, a researcher is called as "node" or "vertex" and the connection is an "edge" represents the co-authorship relationship. If we consider two authors *a1* and *a2* with a publication then the graph can be shown as Fig1. For multiple authors named as *a3*, *b1*, *b2*, *c1* and *c2* if we draw a network using netwokX it will look like Fig1. Research professionals are dedicated to the advancement of the knowledge and practice of professions through developing, supporting, regulating and promoting professional standards for technical and ethical competence. Research professionals' relationship network is represented by undirected weighted graph. This network represents only the co-authorship relationship between research professionals no to how many papers published by one of those authors. Research articles are the product of research professionals' relationship network. The quality of research articles depends on the author who wrote the articles. If an author having relation with other prominent author then have a probability to publish good no. of papers with good quality. In eigenvector centrality, the centrality score of a node is dependent on the number of nodes in neighbors and also the quality of neighbors. So, eigenvector

centrality is more suited for analyzing research professionals' relationship network
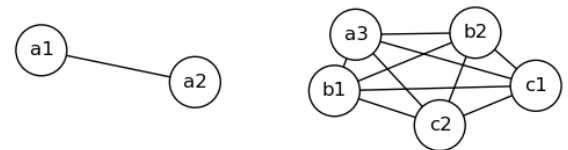

Fig. 1.   Research Professionals' relationship network an example

A. Social Network Analysis (Methodology)

Social Network Analysis (SNA) views social relationships in terms of network theory, consisting of nodes, representing individual actors within the network, and ties which represent relationships between the individuals, such as friendship, kinship, organizations and sexual relationships [2] [5] [8] [10] [11][20]. These networks are often depicted in a social network diagram, where nodes are represented as points and ties are represented as lines. Social network analysis is the mapping and measuring of relationships and flows between people, groups, organizations, computers, URLs, and other connected information/knowledge 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 [13] [17].

In general, the benefit of analyzing social networks is that it can help people to understand how to share professional knowledge in an efficient way and to evaluate the performance of individuals, groups, or the entire social network [2].

B. Eigenvector Centrality

Eigenvector centrality is a measure of the influence of a node in a network. Eigenvector centrality is extension of degree centrality [15]. It is the extension of degree centrality. In degree centrality, the degree centrality of a node is simply count the total no of node that are connected, but in eigenvector centrality, not only consider the total no of adjacent nodes also consider the importance of the adjacent node. In eigenvector centrality, all connections are not equal. In general, connections with influenced person will lend a person more influence than connection with less influenced persons. In eigenvector centrality not only the connections are

important also the score (eigenvector centrality) of the connected node. Eigenvector centrality is calculated by assessing how well connected an individual's to the parts of the network with the greatest connectivity. Individuals with high eigenvector scores have many connections, and their connections have many connections, and their connections have many connections out to the end of the network. Eigenvector centrality is simply dominant of eigenvector of the adjacency matrix. Philip Bonacich proposed eigenvector centrality in 1987 and Google's PageRank is a variant of it [3] [7] [4].

Eigenvector centrality based on adjacency matrix : For a given graph G:=(V,E) with |V| number of vertices. Let A = ($a_{v,t}$) be the adjacency matrix of a graph G is the square matrix with rows and columns labeled by the vertices and entries [4] [16]. i.e.

$$a_{v,t} = \begin{cases} 1, & if\ vertex\ is\ linked \\ 0, & otherwise \end{cases} \qquad (1)$$

If we denote the centrality score of vertex v by X_v, then we can allow for this effect by making X_v proportional to the average of the centralities of v's network neighbors:

$$X_v = \frac{1}{\lambda}\sum_{t \in M(v)} X_t = \frac{1}{\lambda}\sum_{t \in G} a_{v,t} X_t \qquad (2)$$

Where M(v) is a set of the neighbors of v and λ is a constant. With a small rearrangement this can be written in vector notation as the Eigenvector equation

$$\qquad (3)$$

Hence we see that X is an eigenvector of the adjacency matrix of A with eigenvalue λ and λ must be the largest eigenvalue (using the Perron-Frobenius theorem) of adjacency matrix A and X is the corresponding eigenvector [4] [16]. The eigenvector centrality defines in this way accords each node centrality depends on both numbers of connection and quality of its connection. The eigenvector centrality of a node is high either it has large number of connections or it has less number of connections with high scoring node.

The gist of the eigenvector centrality is to compute the centrality of node as a function of the centralities of its neighbors [4].

Eigenvector and Eigenvalue: An Eigenvector of a square matrix A is a non-zero vector v that when the matrix is multiplied by v, yields a constant multiple of v, the multiplier being commonly denoted by λ . That is

$$Av = \lambda v \qquad (4)$$

The number λ is called eigenvalue of *A* corresponding to v. High eigenvector centrality individuals are leaders of the network. They are often public figures with many connections to other high-profile individuals. Thus, they often play roles of key opinion leaders. A related example of this is Google's page rank algorithm, which is closely related to eigenvector centrality calculated on websites based on links to them. High eigenvector centrality individuals, however, cannot necessarily perform the roles of high closeness and betweenness. They do not always have the greatest local influence and may have limited brokering potential.

C. Research Professionals' Relationship Network

In recent years, there have been a sharp increasing number of relationships between research professionals. By jointly publishing a paper, researchers show their knowledge sharing activities, which are essential for knowledge creation. In recent year most research works are done in the collaborative way because research projects are too large for individual researchers, so these often need collaboration from other researchers' which can be same area's researcher or different area's researcher. An important result of research professionals' relationship network is the creation of new scientific/research knowledge, exchange knowledge, finding new research query, new publications, new invention, more number of articles publication. The products of research professionals' relationship are papers. These papers will impact the researchers who are read these articles, and researchers may found new ideas and express these ideas in an article. It is a cyclic process of spreading knowledge and innovation in research community.

II. DATA COLLECTION, CLEANSING AND ANALYSIS

In the process of publishing articles in journals and conferences, publishing and editing books, knowledge does not only exist in a particular researcher's mind but also kept in the papers. Currently, it is not clear which publication data is useful for evaluating the research community. Although, there is a large set of potential data like joined conference organization, joined research proposal submissions, joined publications, joined conference attendance and teacher-student relationship and different sources like IEEE Xplore, Google scholar, DBLP etc. are available. For this analysis, we only considered joined publication as a measure and the data source is IEEE Xplore. IEEE Xplore [1] provides the facility to search articles using various dimensions like date, publication type etc. The search result can be downloaded into CSV (Comma Separated Values) format. We search for articles detail topic wise for the period Jan-2000 to Jan-2014 and export the search result. Exported search result is in CSV format. The CSV file has 33 fields. The field names are as follows, "Document Title, Authors, Author Affiliations, Publication Title, Publication Date, Publication Year, Volume, Issue, Start Page, End Page, Abstract, ISSN, ISBN, EISBN, DOI, PDF Link, Author Keywords, IEEE Terms, INSPEC Controlled Terms, INSPEC Non-Controlled Terms, DOE Terms, PACS Terms, MeSH Terms, Article Citation Count, Patent Citation Count, Reference Count, Copyright Year, Online Date, Date Added To Xplore, Meeting Date, Publisher,

Sponsors, Document Identifier". We have filtered the data and select only Document Title, Authors, Publication Title, Article Citation Count, Publisher and Document Identifier. Authors' name of an articles are stored in a single column which is separated by a ";". Author and its co-author are extracted based on delimiter ";" and store each author name in individual column.

After extraction of author's name we found that some of author's names are unable to read, so data cleaning has become necessary to clean such type of names by replacing actual names. We omitted the name of authors and marked each of them a personal number from 1 to 61546. After cleaning of publication data, 26802 publication and 61546 authors were finally available for analysis. We found that a paper has been written by minimum one author and maximum of 59 authors. Averagely each of these papers is written around two to four authors. In the dataset there are two papers that were composed by 59 authors which has the most co-published paper and 3230 papers is written by only one authors which has the least co-published paper. The most prolific writer is Lau, Y.Y., who published 54 papers. Author Sasaki, M. works with maximum number of co-authors is 63 and some of author works individually or have one or two co-authors.

A. Co-authorship network of research professionals'

Research professionals' relationship network is drawn based on co-author; means an author have connection with those authors who have published articles in journal and conference together, publish and edit books together, publish articles in transaction together which can indicate individuals' status and the scientific collaboration. Based on the available publication data of researcher, we can build a network matrix which is representing the relationship between researchers according to following technique: For instance, there are three papers: P1, P2 and P3. P1 and P2 are conference paper and P3 is journal. P1 has two authors' a_1 and a_2 and citation is 12, P2 has three authors' b_1, a_3 and b_2 and citation is 2. P3 has five authors' c_1, a_3, b_2, c_2 and b_1 and citation is 11. It shows like this: P1-{a_1, a_2}, P2-{b_1, a_3, b_2}, P3-{c_1, a_3, b_2, c_2, b_1}[14]. After that we extract author and its co-author who are involved in their research work and calculate the collaboration weight. Author and co-author are {a_1- a_2}, {b_1- a_3, b_1 - b_2, a_3 - b_2}, {c_1 - a_3, c_1 - b_2, c_1 - c_2, c_1 - b_1, a_3 - b_2, a_3 - c_2, a_3 - b_1, b_2 - c_2, b_2 - b_1, c_2 - b_1} [9] [11] [12] [18] [19] and calculate the relationship weight based on the following parameter:

- Simple adjacency weight means relationship weight is 1.

- Total number of publication that are published together.

- Total citation count of articles that are published together.

- Priority weight of articles. Here we set the priority weight of articles which is published in conference and journals. We set the priority weight of conference is 1 and journal is 2. After that we calculate the total priority

weight of articles that are published in conference and journals together.

Then we generate four different networks based on different relationship weight by using python and networkX [5] of these data.

B. Analysis and Result

Eigenvector centrality is the extension of degree centrality, in degree centrality which simply counts the total number of node those are adjacent to a node. But eigenvector centrality not only considers the number of connections also quality of connections. Research professionals' relationship network is a co-authorship relationship network. The product of research professionals' relationship network is research articles and quality of research articles are depends on authors of articles, if the authors are research professionals and they have relation with others research professionals then have a probability to publish a good quality and quantity of papers than a researcher have relation with same no. of research scholar or students. Therefore, we can say that eigenvector centrality is suitable for finding the key author or the community in the network. To compute eigenvector centrality is to give starting random positive amount of influence of every node and then use the power method with the largest eigenvalue of the adjacency matrix of graph. Here we use simple and weighted both types of adjacency matrix for eigenvector centrality calculation. Weights are total no. of publications, total citation count and priority weight. Now, we calculate eigenvector centrality for all research professionals by using python and networkx. After that we export all the results in MYSQL database and arrange the authors in descending order, then select top 10 authors from each measures and combined it and obtain 26 important authors shown in table I, in which the sequence of the authors are listed firstly by the decreasing order of adjacency weight and then by frequency weight, citation count and priority weighted eigenvector centrality of authors. These 26 authors' data are available for analysis.

TABLE I.   TOP 26 RESEARCHER ACCORDING TO DIFFERENT WEIGHT MEASURE IN THE NETWORK

Sl.	Author Id	Eigenvector Centrality Value (Adj)	Eigenvector Centrality Value (Fre)	Eigenvector Centrality Value (cit)	Eigenvector Centrality Value (wt)
1	48688	0.1334739264	0.1319292781	0.0000000000	0.1586533880
2	5340	0.1333693085	0.1530856888	0.0000000000	0.1586398476
3	56856	0.1295166699	0.1206014135	0.0000000000	0.1291898487
4	36605	0.1288715982	0.1220020498	0.0000000000	0.1525935228
5	48451	0.1284210454	0.1681615443	0.0000000000	0.1493243099
6	43010	0.1284210454	0.1582503275	0.0000000000	0.1493243099
7	45262	0.1284210454	0.1534493371	0.0000000000	0.1493243099
8	29619	0.1284210454	0.1506036831	0.0000000000	0.1493243099
9	20758	0.1284210454	0.1504937042	0.0000000000	0.1493243099
10	3022	0.1284210454	0.1473571172	0.0000000000	0.1567190744
11	22469	0.1284210454	0.1473497239	0.0000000000	0.1493243099

12	17849	0.1284210454	0.1464638712	0.0000000000	0.1567190744
13	23930	0.1284210454	0.1414998703	0.0000000000	0.1493243099
14	1370	0.1284210454	0.1374289909	0.0000000000	0.1493243099
15	11735	0.1284210454	0.1195723732	0.0000000000	0.1493243099
16	34608	0.1235593450	0.1464513627	0.0000000000	0.1194469230
17	17718	0.0000000002	0.0000000005	0.0079086424	0.0000000001
18	16409	0.0000000000	0.0000000000	0.0079072820	0.0000000000
19	33284	0.0000000000	0.0000000000	0.0079072820	0.0000000000
20	27323	0.0000000000	0.0000000000	0.0025384326	0.0000000000
21	11029	0.0000000000	0.0000000019	0.5771952205	0.0000000016
22	189	0.0000000000	0.0000000001	0.5773567438	0.0000000001
23	18993	0.0000000000	0.0000000001	0.5771597895	0.0000000001
24	18913	0.0000000000	0.0000000000	0.0115045737	0.0000000000
25	14024	0.0000000000	0.0000000000	0.0025384086	0.0000000000
26	18519	0.0000000000	0.0000000000	0.0025384086	0.0000000000

After the analysis we found that no one author having high value in all measure but some of the authors having high value based on adjacency weight, frequency and priority weight. According to adjacency weight author Mizuno, T. (ID=48688) is key author, according to frequency weight author Spandre, G. (ID=48451) becomes the key author, according to citation weight author Candes, E.J.(ID=189) becomes a key author, according to priority weight author Mizuno, T.(ID=48688) becomes a key author. Author Mizuno, T. (ID=48688) is key based on adjacency and priority weight, so we can say that author Mizuno, T.(ID=48688) is key author in the network but in citation weight network the eigenvector centrality value is high than other weighted network.

After the study and analysis of eigenvector centrality algorithm of networkX we found that the eigenvector centrality calculation is done by the power iteration method. Firstly, set starting random positive amount of influence of every node and then use the power method to find the eigenvector for the largest eigenvalue of the adjacency matrix of graph. The iteration will stop after max iterations or an error tolerance has been reached. In power iteration method, at every iteration vector is multiplied by the adjacency matrix of graph and normalized. In networkX eigenvector centrality algorithm, the starting amount of influence of each node is $1/len(G)$. Where $len(G)$ is the total number of nodes in the graph. We are known that the Eigenvector centrality is the extension of degree centrality. Our approach is to compute eigenvector centrality based on degree centrality. Therefore, we set the starting amount of influence of each node as degree centrality of that node, and then use power method to calculate the eigenvector centrality of each node. So here we modify the eigenvector centrality algorithm of networkX and calculate eigenvector centrality of each node. We select top 10 authors from each measures and combined it and obtain 26 important authors shown in table II, in which the sequence of the authors are listed firstly by the decreasing order of adjacency weight and then by frequency weight, citation count and priority weighted eigenvector centrality of authors. These 26 authors' data are available for analysis.

TABLE II. TOP 26 RESEARCHER ACCORDING TO DEGREE CENTRALITY BASED EIGENVECTOR CENTRALITY WITH DIFFERENT WEIGHT MEASURES

Sl..	Author ID	Eigenvector Centrality(Adj)	Eigenvector Centrality (Frequecny)	Eigenvector Centrality (Citation)	Eigenvector Centrality(Priority Weight)
1	48688	0.1334739264	0.1319293541	0.0000000000	0.1586587791
2	5340	0.1333693085	0.1530859655	0.0000000000	0.1586452382
3	56856	0.1295166699	0.1206014175	0.0000000000	0.1291942489
4	36605	0.1288715982	0.1220027343	0.0000000000	0.1525986826
5	48451	0.1284210454	0.1681624648	0.0000000000	0.1493293598
6	43010	0.1284210454	0.1582512672	0.0000000000	0.1493293598
7	45262	0.1284210454	0.1534502347	0.0000000000	0.1493293598
8	29619	0.1284210454	0.1506045342	0.0000000000	0.1493293598
9	20758	0.1284210454	0.1504945628	0.0000000000	0.1493293598
10	3022	0.1284210454	0.1473579923	0.0000000000	0.1567243764
11	22469	0.1284210454	0.1473505661	0.0000000000	0.1493293598
12	17849	0.1284210454	0.1464647160	0.0000000000	0.1567243764
13	23930	0.1284210454	0.1415006600	0.0000000000	0.1493293598
14	1370	0.1284210454	0.1374297522	0.0000000000	0.1493293598
15	11735	0.1284210454	0.1195730779	0.0000000000	0.1493293598
16	34608	0.1235593450	0.1464521887	0.0000000000	0.1194509617
17	17718	0.0000000002	0.0000000003	0.0034618814	0.0000000002
18	16409	0.0000000000	0.0000000000	0.0034585634	0.0000000000
19	33284	0.0000000000	0.0000000000	0.0034585634	0.0000000000
20	54375	0.0000000000	0.0000000002	0.0015333183	0.0000000007
21	36033	0.0000000000	0.0000000078	0.0021154099	0.0000000460
22	11029	0.0000000000	0.0000000006	0.5777020637	0.0000000017
23	16359	0.0000000000	0.0000000009	0.0021142049	0.0000000026
24	189	0.0000000000	0.0000000000	0.5772160296	0.0000000001
25	18993	0.0000000000	0.0000000000	0.5769417419	0.0000000001
26	18913	0.0000000000	0.0000000000	0.0115131276	0.0000000000

After analysis we found that no one author having high value in all measure but some of the authors having high value based on adjacency weight, frequency and priority weight. According to adjacency weight author Mizuno, T. (ID=48688)is key author, according to frequency weight author Spandre, G. (ID=48451) becomes the key author, according to citation weight author Romberg, J.(ID=11029) becomes a key author, according to priority weight author Mizuno, T.(ID=48688) becomes a key author. Author Mizuno, T. is key based on adjacency and priority weight, so we can say that author Mizuno, T. is key author in the network but in citation weight network the eigenvector centrality value is high than other weighted network.

III. CONCLUSION

In this paper we discuss eigenvector centrality and its application. We calculate eigenvector centrality of each node with different weight and also calculate eigenvector centrality

based on degree centrality of each research professionals in the research professionals' relationship network. We found that the results of eigenvector centrality are different on different weight measures in both traditional eigenvector centrality and degree centrality based eigenvector centrality and no one author having high value in all measures. Degree centrality based eigenvector centrality in citation weighted network gives better result than others.

REFERENCES

[1] "http://ieeexplore.ieee.org/xpl/opac.jsp."

[2] A. Abbasi and J. Altmann, "On the correlation between research performance and social network analysis measures applied to research collaboration networks," in 44th Hawaii International Conference on System Sciences (HICSS), 2011. IEEE, 2011, pp. 1–10.

[3] P. Bonacich and P. Lloyd, "Eigenvector-like measures of centrality for asymmetric relations," Social Networks, vol. 23, no. 3, pp. 191–201,2001.

[4] S. A. Burr, *The mathematics of networks*. American Mathematical Soc., 1982, no. 26.

[5] C. Correa, T. Crnovrsanin, and K.-L. Ma, "Visual reasoning about social networks using centrality sensitivity," IEEE Transactions on Visualization and Computer Graphics,, vol. 18, no. 1, pp. 106–120, 2012.

[6] P. J. S. D. A. S. Aric A. Hagberg, "Exploring network structure, dynamics,and function using networkx," proceedings of the 7th python in science conference (scipy 2008), 2008.

[7] D.-w. Ding and X.-q. He, "Application of eigenvector centrality in metabolic networks," in 2nd International Conference on Computer Engineering and Technology (ICCET), 2010, vol. 1. IEEE, 2010, pp. V1–89.

[8] D.-M. B. Friedl, J. Heidemann et al., "A critical review of centrality measures in social networks," Business & Information Systems Engineering, vol. 2, no. 6, pp. 371–385, 2010.

[9] Y. He and S. Cheung Hui, "Mining a web citation database for author co-citation analysis," Information processing & management, vol. 38, no. 4, pp. 491–508, 2002.

[10] B. Liu, Web data mining. Springer, 2007.

[11] X. Liu, J. Bollen, M. L. Nelson, and H. Van de Sompel, "Co-authorship networks in the digital library research community," Information processing & management, vol. 41, no. 6, pp. 1462–1480, 2005.

[12] M. E. J. Newman, "Coauthorship networks and patterns of scientific collaboration," Proceedings of the National Academy of Sciences, vol. 101, no. suppl 1, pp. 5200–5205, 2004.

[13] Y. H. Said, E. J. Wegman, W. K. Sharabati, and J. T. Rigsby, "Retracted: Social networks of author–coauthor relationships," Computational Statistics & Data Analysis, vol. 52, no. 4, pp. 2177–2184, 2008.

[14] Y. Said, E. Wegman, W. Sharabati, and J. Rigsby, "Social networks of author-coauthor relationships (retraction of vol 52, pg 2177, 2008)," COMPUTATIONAL STATISTICS & DATA ANALYSIS, vol. 55, no. 12, pp. 3386–3386, 2011.

[15] L. Spizzirri, "Justification and application of eigenvector centrality."

[16] P. D. Straffin, "Linear algebra in geography: Eigenvectors of networks," Mathematics Magazine, pp. 269–276, 1980.

[17] J. Tang, D. Zhang, and L. Yao, "Social network extraction of academic researchers," in Seventh IEEE International Conference on Data Mining, 2007. ICDM 2007. IEEE, 2007, pp. 292–301.

[18] B. Wang and J. Yang, "To form a smaller world in the research realm of hierarchical decision models," in Proc. PICMET'11. PICMET, 2011.

[19] B. Wang and X. Yao, "To form a smaller world in the research realm of hierarchical decision models," in International Conference on Industrial Engineering and Engineering Management (IEEM), 2011 IEEE. IEEE, 2011, pp. 1784–1788.

[20] S. Wasserman, *Social network analysis: Methods and applications*. Cambridge university press, 1994, vol. 8.