Network and Complex Systems
ISSN 2224-610X (Paper) ISSN 2225-0603 (Online)
Vol.4, No.5, 2014
www.iiste.org

A new Approach to Erdòs Collaboration Network using PageRank

Weijia Qian, Feifan Zhou, Chengyuan Guan, Junwei Ding, Shen Zhang, Chuan Shang, Huiling Cao*
Aeronautical Engineering Institute, Civil Aviation University of China, Tianjin, 300300, China
* E-mail of the corresponding author: hlcao@cauc.edu.cn

Abstract
Using the data on Paul Erdòs, his co-authors and their co-authors, we can construct a network called the Erdòs Collaboration network. Then we do reduction, analysis and visualization with it using program Pajek. In this paper, we develop a reasonable academic influence measuring method applying PageRank algorithm on the case of the Erdòs Collaboration network. We find that ALON, NOGA M is the most influential mathematician in the network. In addition, to measure impact, we construct a dynamic model, whereas it needs too much data for us to calculate the dynamic index.
Keywords: PageRank, Collaboration network, Network analysis.

1. Introduction

1.1 Existing Methods of Measuring Academic Influence

It is necessary to measure the academic performance of a given researcher in many situations such as allocating research grants, recruiting and re-appointing. An easy way is to count a given researcher's published papers and the sum of *citations count* of his papers, where *citations count* is " how many times the published paper is cited by other papers"[Abbas 2011]. In our opinion, this way of counting citations cannot accurately reflect the level of a paper, let alone a researcher.

Another measuring method called "h-index" is proposed by Jorge Hirsch [2005]. Although "h-index" is deemed to be a good measure by many researchers [Healy et al. 2011], many improved methods are proposed by others [Anderson et al., 2008; Egghe, 2006].

While two methods above can reflect the academic ability "quantitatively", they surely ignore cooperation which is an important element in today's academic career. Hence, we construct a network-based model for calculating academic influence derived from cooperation and try to apply the PageRank algorithm.

1.2 Erdòs Collaboration Graph

Paul Erdòs is a legendary mathematician who cooperated with over 600 mathematicians and published more than 1,500 papers. He won many prizes including Cole Prize in 1951 and the Wolf Prize in1983. He is also the mathematician having the largest number of different co-authors and a promoter of collaboration.

The Erdòs number of a mathematician is defined as follows: Paul Erdòs has a Erdòs number of 0; people who have co-written a paper with Erdòs have a Erdòs number of 1 and their co-authors have a Erdòs number of 2; etc.

The data about Erdòs's co-authors and their co-authors can be represented in the Erdòs collaboration graph C which sees mathematicians as its vertices and describes cooperation within them using edges i.e., there is an edge between two vertices if they have published at least one paper together [Oakland University, 2014]. The mathematicians in the graph all have a Erdòs number varying from 0 to 2. For simplification, we do not include Erdòs, so we can delete that vertex and achieve a "truncated Erdòs collaboration graph C'" [Batagelj and Mrvar, 2000].

We can use program Pajek to analyse the graph. Program Pajek is a powerful program for big networks analysis and visualization [Batagelj and Mrvar, 1998] and it is free for noncommercial use.

2. Construction and Analysis of the Erdos Collabopation Network

2.1 Construction

In brief, the construction of the network consists of steps as follows: first, collect data; then determine all vertices and links; at last, the the value of vertices and links. The network $N=(V,L,P,W)$ that we want to build consists of four parts:

- V is the set of vertices. Each element of V represents a mathematician whose Erdòs number is of 1 or 2 .
- L is the set of links. In this network, all the links are undirected. If two mathematicians publish at least one paper once together, then we add a link between them.

- P is the vertices values functions.
- W is the edges values functions.

*vertices *n*	the network has n vertices
1 "name 1"	
2 "name 2"	the label for vertices n is "name n"
...	
n "name n"	
*arcs	
*edges	
i j v_{ij}	
...	the edge between i and j has value V_{ij}

Figure 1.The Net Format for Pajek.

At the beginning, we need to translate the html data into net data (supported by program Pajek) with a Matlab program.
The network in net format is of the form as Figure 1 Shows, but always much more larger [Batagelj, 2002]. After loading the data, we could do analysis and visualization using Pajek.

2.2 General Information and Visualization

First of all, we ought to reduce the network, because vertices with degree 1 do not require consideration. Then, we run Pajek and it outputs the general information of the network (**Table 1**) and finishes visualization (**Figure 2**).

2.3 Network Properties

- **Degree:** We firstly achieve the distribution of degrees of mathematicians (**Figure 3**). The largest degree is 200 of ALON, NOGA M.
- **Diameter:** It is defined as the longest shortest path in the network, and it is from FOWLER, THOMAS GEORGE to OFFORD, ALBERT CYRIL*. So the diameter of this network is 8.
- **Vertices:** The network that we analyze has been reduced, so it has only one component which contains 3,221 vertices.
- **Core:** Seidman [1983] introduced the concept of core. In the reduced network, the main core is of order 12.

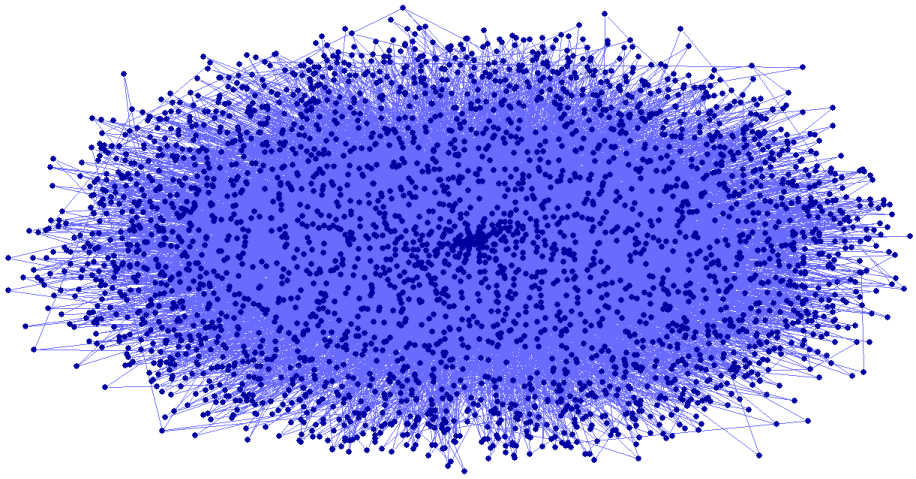

Figure 2.Visualization of the Reduced Co-author Network.
It has only one component due to reduction.

3. PageRank Algorithm Application

3.1Assumptions

- The academic influence of a researcher is proportional to the network importance of a vertex .
- The larger a vertex degree is, the more influential a vertex is.

- If a vertex is linked to an influential adjacent vertex, then we can consider it more "influential".

3.2PageRank Algorithm

It is really hard to present a precise definition of influence, but we can say that the more influential a researcher is, there are more peers who want to collaborate with him. In addition, in the case that two researchers have the same collaboration, the researcher whose co-authors are more influential should be considered more influential. It inspires us with **PageRank algorithm** which is invented by Larry Page and Sergey Brin [Page and Brin, 1998]. Let I_p denote the influence rank originated from PageRank.

First, we should produce a link matrix $\boldsymbol{L}$ using data of edges in the reduced network exported from Pajek, where

$$l_{ij} = \begin{cases} 1, & \text{if i and j are linked} \\ 0, & \text{otherwise} \end{cases}.$$

Because the reduced network is undirected, the link matrix is symmetric. We normalize all rows and achieve a link transition probability matrix **Q**. Then, we amend it to get Google matrix **G**,

$$\mathbf{G} = \alpha \cdot \mathbf{Q} + (1-\alpha) \cdot (1/\mathrm{N}) \cdot \mathbf{e} \cdot \mathbf{e}^{\mathrm{T}},$$

where $\alpha = 0.85$, N=3221 and $e = \left(\underbrace{1,1,\ldots,1,1}_{3321} \right)^T$. At last, we calculate the eigenvector of eigenvalue 1 with Matlab, and I_p equals the corresponding element in the eigenvector.

4. Results and Discussion

Top 15 influential researchers according to PageRank are list in Table 2. We also provide some measures like vertex degree , total citation for comparison in Table 2. According to our model, we can find ALON, NOGA M is the most influential researcher who has 200 co-authors and citation of 15152 and co-writes 5 papers with Erdös. The method of measuring academic influence with PageRank is reasonable, while it should be incorporated with other measures such as core number.

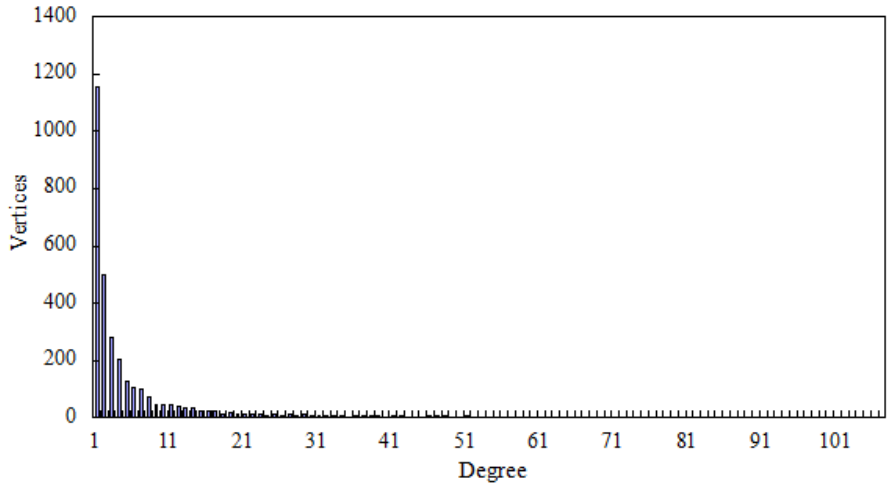

Figure 3.The Distribution of Degree.

This is a typical power-law graph, which is an important property of the co-author network.

5. Outlook: Measuring Instantaneous Influence (IMPACT)

We can define the instantaneous influence as the impact of a vertex. We develop a dynamic measure model

$$\mathrm{Impact}(t) = \frac{\Delta \mathrm{I}(t)}{\Delta t},$$

where $\Delta \mathrm{I}(t)$ is the variation of an influence measure and $\Delta \mathrm{t}$ is the time step which should be determined according to specific problems. This impact measure can detect the vertex which produces a huge "impact" at time *t*. But the calculation of this impact measure needs abundant historical data which propose a hinder for us.

Table 1.General information about the network.

	Vertices	Edges	Density	Average degree
Before reduction	10242	16981	0.000324	3.32
After reduction	3221	14670	0.00283	9.11

We add the general information before reduction for comparison. Thus we can see the effects of reduction.

Table 2.Results: top 15 researchers corresponding to PageRank.

Rank	Mathematician	Degree	Citation	PageRank
10	ALON, NOGA M.	200	28437	0.412447159
82	COLBOURN, CHARLES JOSEPH	191	4444	0.373339108
41	BOLLOBAS, BELA :	141	4840	0.261652067
58	CAMERON, PETER J.	127	3010	0.250862872
74	CHUNG, FAN RONG KING (GRAHAM) :	123	7498	0.224164158
15	ARONOV, BORIS	97	1840	0.203014617
73	CHUI, CHARLES KAM-TAI	87	4231	0.190796776
19	BABAI, LASZLO	95	5213	0.185845971
65	CHARTRAND, GARY THEODORE	106	938	0.18022785
2	ACZEL, JANOS D.	77	1215	0.1606561
18	AVIS, DAVID MICHAEL	78	2228	0.15266747
67	CHEN, GUANTAO	72	322	0.143358338
2390	TUZA, ZSOLT	102	1584	0.033185535
1347	NESETRIL, JAROSLAV	73	1421	0.023390802
2844	Wilf, Herbert S.	77	2090	0.023168763

References

Abbas, Ash Mohammad. "Weighted indices for evaluating the quality of research with multiple authorship." *Scientometrics* 88.1 (2011): 107-131.

Hirsch, Jorge E. "An index to quantify an individual's scientific research output." *Proceedings of the National academy of Sciences of the United States of America* 102.46 (2005): 16569-16572.

Healy, N. A., et al. "The h index and the identification of global benchmarks for breast cancer research output." *Breast cancer research and treatment*127.3 (2011): 845-851.

Anderson, Thomas R., Robin KS Hankin, and Peter D. Killworth. "Beyond the Durfee square: Enhancing the h-index to score total publication output."*Scientometrics* 76.3 (2008): 577-588.

Egghe, Leo. "An improvement of the H-index: the G-index." *ISSI newsletter*2.1 (2006): 8-9.

Oakland University.Information about the Erdös Number Project - The Erdös Number Project.(2014) http://www.oakland.edu/enp/readme/

Batagelj, Vladimir, and Andrej Mrvar. "Some analyses of Erdos collaboration graph." *Social Networks* 22.2 (2000): 173-186.

Batagelj, Vladimir, and Andrej Mrvar. "Pajek — a program for large network analysis". (1998)

Seidman, Stephen B. "Network structure and minimum degree." *Social networks* 5.3 (1983): 269-287.

Page, Lawrence, et al. "The PageRank citation ranking: Bringing order to the web." (1999).

The IISTE is a pioneer in the Open-Access hosting service and academic event management. The aim of the firm is Accelerating Global Knowledge Sharing.

More information about the firm can be found on the homepage:
http://www.iiste.org

CALL FOR JOURNAL PAPERS

There are more than 30 peer-reviewed academic journals hosted under the hosting platform.

Prospective authors of journals can find the submission instruction on the following page: http://www.iiste.org/journals/ All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Paper version of the journals is also available upon request of readers and authors.

MORE RESOURCES

Book publication information: http://www.iiste.org/book/

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar