A new Approach to Erdős 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 Erdős, his co-authors and their co-authors, we can construct a network called the Erdős 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 Erdős 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 Erdős Collaboration Graph
Paul Erdős is a legendary mathematician who cooperated with over 600 mathematicians and published more than 1,500 paper. He won many prizes including Cole Prize in 1951 and the Wolf Prize in 1983. He is also the mathematician having the largest number of different co-authors and a promoter of collaboration.

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

The data about Erdős’s co-authors and their co-authors can be represented in the Erdős 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 Erdős number varying from 0 to 2. For simplification, we do not include Erdős, so we can delete that vertex and achieve a “truncated Erdős 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 Erdős Collaboration 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 Erdős 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.
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.

3. PageRank Algorithm Application

3.1 Assumptions

- 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”.

3.2 PageRank 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 the 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 \( L \) using data of edges in the reduced network exported from Pajek, where

\[
L_{ij} = \begin{cases} 
1, & \text{if } i \text{ and } j \text{ are linked} \\
0, & \text{otherwise}
\end{cases}
\]

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 \( G \),

\[
G = \alpha \cdot Q + (1 - \alpha) \cdot (1/N) \cdot e \cdot e^T,
\]

where \( \alpha = 0.85 \), \( N = 3221 \) and \( e = \left[ 1,1,\ldots,1,1 \right]_{3221} \). 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 listed 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 Erdős.

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.](image)

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

\[
\text{Impact}(t) = \frac{\Delta I(t)}{\Delta t},
\]

where \( \Delta I(t) \) is the variation of an influence measure and \( \Delta t \) 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>
<thead>
<tr>
<th>Table 1. General information about the network.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vertices</strong></td>
</tr>
<tr>
<td>Before reduction</td>
</tr>
<tr>
<td>After reduction</td>
</tr>
</tbody>
</table>

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.

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

References


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


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](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/](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**


**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 Digital Library, NewJour, Google Scholar