Determination of the Influence of Academic Research Base on Social network analysis

Zhihui Yang[a*], Cunbin Zhu[b]

School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu 233030, China

[a] email: zhi-hui-yang@163.com, [b]email: dw_zcb@163.com

* the corresponding author

Keywords: Social network analysis; Networks' properties; Measure of Influence

Abstract. One way to determine the influence of academic research is to build co-author networks. Co-authoring usually means strong link between researchers. A typical example is Paul Erdös. We have to discuss the following problem. In this paper, we should build a co-author networks from the file Erdos1, which includes 511 researchers but do not include Erdös. Then we obtain the correct set of nodes (the Erdös coauthors) and their links (connections with one another as co-authors). We need calibrate our model by limiting the size of the co-author networks. Once built, we analyze the properties of this network, which includes Problem 2 mentioned that selecting the most influential researchers in file Erdos1.

Introduction

Interdisciplinary cooperation becomes more and more popular today, Co-author is one form. The Researches on co-author networks across the world covers the following aspects. Xiaoming Liu, Johan Bollen studied the properties of co-author networks in digital library. They proposed algorithm that can be used to calculate the weighted network centrality, which is also called Author Rank algorithm. And this algorithm stems from the Pagerank algorithm [1]. Hou H analyzed the data from the published literature between 1978 and 2004 in the SCI. Through social network analysis, symbiotic analysis, clustering analysis and routine analysis in his study, he found the microstructure of co-author networks in scientific measure, the main areas of the whole network and different cooperation network, and the central cooperation network during the investigation [2]. On the type of Co-authoring, Newman ME J. studied three areas of physics, biomedical and computer science by comparing the co-author networks. He pointed out the differences between the types of disciplines coauthored [3]. On the Cooperation strength, Kretschmer mainly researched on the relationship between scientific field and structural characteristics of the network. He studied the co-author networks in information visualization field. He analyzed the evolution of Co-author networks, the author strength, Degree and strength distribution, Sub-groups size and distribution, the homogeneity of the intensity distribution [4].

Key Terminologies and Assumptions

Degree. The degree of a node is the number of links that involve that node, which is the cardinality of its neighborhood [5].

Centrality. In graph theory and network analysis, centrality of a vertex measures its relative importance within a graph. Applications include how influential a person is within a social network, how important a room is within a building (space syntax), and how well-used a road is within an urban network. There are four main measures of centrality: degree, betweenness, closeness, and eigenvector. Centrality concepts were first developed in social network analysis, and many of the terms used to measure centrality reflect their sociological origin [6].

Density. The number of ties in the ego network (not counting ties involving ego) divided by the number of pairs of alters in the ego network (i.e., potential ties).

Clustering Coefficient. If author m is one of authors of a co-authorship network, the clustering coefficient is the probability that one of co-authors of m co-author with m's another co-author. The clustering coefficient of any author k can be calculated by Petter Holme's method [7].

Let us formalize the assumptions which shall go into our models.

- Ignore the time factor. The year of joint publications does not affect the model analysis.
- No repeat cooperation papers. We always think there is one cooperation paper if there more than one.
- In the co-author networks model, we just consider the co-authorship factor and ignore other factors.

Analysis of Co-author Networks' properties

We use MATLAB software to choose 511 researchers from the file Erdos1: https://files.oakland.edu/users/grossman/enp/Erdos1.html, who coauthored a paper with Erdös, but do not include Erdös. We should transform their co-authorships to 0-1 matrix based on the available publication data by binary processing. In addition, the 511 authors are marked as number 1~511 in alphabetical orders. After importing the network matrix as a table into UCINET, we calculated the social network measures of our model.

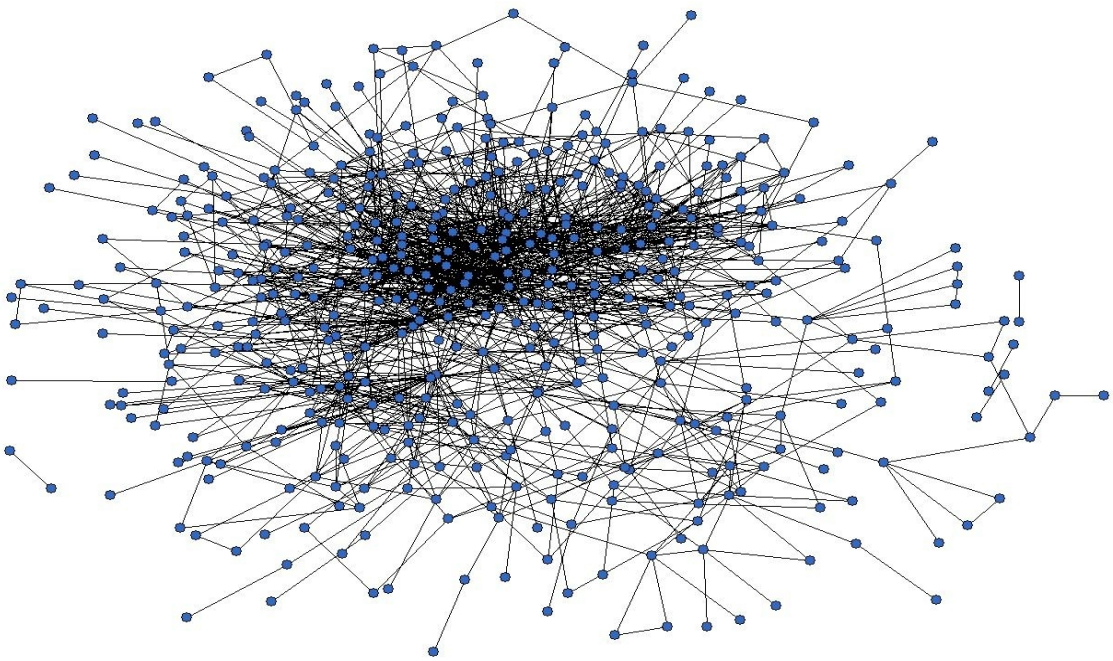

Figure 1. collaboration network structure

Density. Define M_1 equals to actual total of relationship, and M_2 equals to theory total of relationship. Thus

$$M_2 = C_n^2 = n^2 = \frac{511\times 510}{2\times 1} = 130305\,,\ \mathrm{De} = \frac{M_1}{M_2}$$

by use UCINET to reach the solution as De= 0.0125, which stands for the tightness between those authors.

Average Path Length. Distance between actors of network determined the connectivity of network, lower the distance, steadier the networks are. First, we get the distance l_{ij} from i to j. Thus we can calculate the average distance of all nodes:

$$L = \frac{2}{n(\mathrm{n}-1)}\sum_{i<j} l_{ij}$$

and we get the average distance L=3.822.

Clustering Coefficient. In undirected networks, the clustering coefficient C_i of a node i is defined as

$$C_i = \frac{2e_i}{m_i(\mathrm{m_i} - 1)}$$

where m_i is the number of neighbors of n and e_i is the number of connected pairs between all neighbors of n. The network clustering coefficient is the average of the clustering coefficients for all nodes in the network. Thus we can calculate:

$$C = \frac{1}{n}\sum_{i=1}^{n} C_i$$

By use UCINET reach the solution as C= 0.341. It indicates that probability that two authors co-author with the same one is 34.1%

Measure(s) of Influence

We choose all nodes except ones' degree is zero and count frequency of their degree. Then we can gain the probability density as follows:

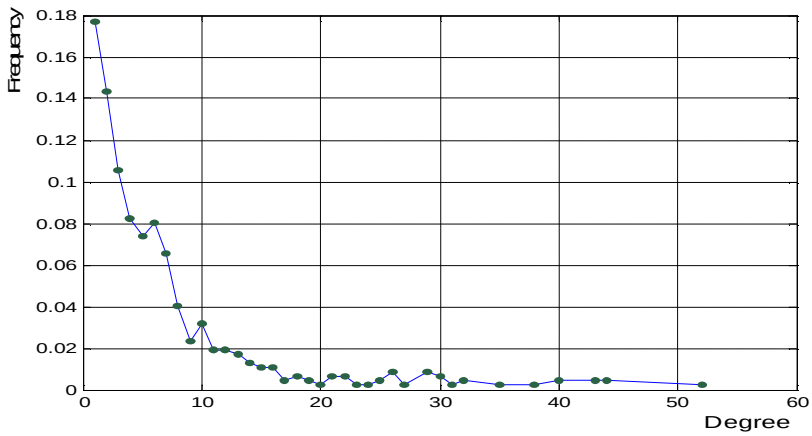

Figure 2. Degree Distribution

As in Fig. 2, degrees approximately subject to Poisson distribution. That is to say, most authors only co-author with fewer others. There are several authors co-authoring with numerous. It is this small part becoming the central nodes of the network, plays a vital role in the network.

Betweenness. We can use UCINET to reach the solution as shown in Table 1:

Table 1. Betweenness

mark	name	Betweenness	nBetweenness
187	HARARY, FRANK*	18899.72	7.281
438	HARARY, FRANK*	17816.18	6.863
385	RUBEL, LEE ALBERT*	17077.76	6.579
449	STRAUS, ERNST GABOR*	16587.79	6.39
355	POMERANCE, CARL BERNARD	15239.81	5.871
148	FUREDI, ZOLTAN	14261.72	5.494
10	ALON, NOGA M.	13886.22	5.349
165	GRAHAM, RONALD LEWIS	13679.02	5.269
44	BOLLOBAS, BELA	13419.08	5.169
341	PACH, JANOS	12104.3	4.663

In order to further understand the influence of projection on author's betweenness, the ranks of authors with top10 largest betweenness values are shown in Table 1. Higher betweenness locate at more linked paths of co-author network. And the nodes such as *HARARY, FRANK** can influent the

group by controlling or distorting delivery of information. Thus these authors are in an important position and possess lager power.

Centrality analysis and consensus analysis. By use UCINET and gain OutDeg, Indeg, OutBonP, InBonPw, Out2Ste, In2Step, OutARD, InARD, Between—all are the abbreviation of parameters. Then we get all nodes K-core and order by it. Select 34 nodes while the K-core is over 8 as the followed table. If we order these nodes by different parameters, we could get different result.

Table 2. K-core

	OutDeg	Indeg	OutBonP	InBonPw	Out2Ste	In2Step	OutARD	InARD	Between
10	0.102	0.1001	31.781	31.311	0.482	0.48	0.421	0.417	0.053
21	0.051	0.051	65.818	65.657	0.324	0.322	0.362	0.359	0.014
44	0.084	0.0841	7.5101	7.216	0.473	0.469	0.411	0.408	0.052
58	0.037	0.037	56.038	55.852	0.329	0.325	0.355	0.352	0.008
78	0.061	0.059	75.65	75.298	0.388	0.384	0.378	0.374	0.022
128	0.059	0.059	83.442	83.083	0.394	0.39	0.381	0.378	0.021
140	0.043	0.043	62.75	62.489	0.355	0.353	0.367	0.365	0.009
148	0.078	0.076	103.535	101.478	0.457	0.439	0.408	0.401	0.055
164	0.035	0.035	52.808	52.501	0.253	0.247	0.333	0.329	0.003
165	0.086	0.086	104.95	104.582	0.484	0.48	0.413	0.411	0.053
177	0.063	0.063	90.904	90.505	0.341	0.339	0.37	0.367	0.016
178	0.025	0.025	48.188	48.307	0.29	0.288	0.341	0.339	0.002
180	0.059	0.059	59.041	58.79	0.361	0.359	0.375	0.373	0.045
187	0.086	0.084	65.063	64.811	0.404	0.4	0.394	0.39	0.073
227	0.039	0.039	45.361	45.165	0.245	0.243	0.332	0.329	0.005
249	0.057	0.057	60.622	60.321	0.39	0.386	0.38	0.377	0.03
261	0.049	0.047	73.293	72.93	0.308	0.306	0.357	0.353	0.006
275	0.041	0.041	65.89	65.567	0.292	0.288	0.348	0.345	0.004
287	0.059	0.057	79.519	79.485	0.41	0.408	0.387	0.383	0.022
290	0.049	0.049	70.525	70.332	0.327	0.325	0.358	0.356	0.01
326	0.057	0.057	75.752	75.491	0.406	0.402	0.378	0.375	0.02
341	0.063	0.063	77.629	77.336	0.39	0.386	0.383	0.379	0.047
363	0.024	0.024	41.633	41.524	0.247	0.245	0.329	0.327	0.002
378	0.084	0.082	118.811	118.335	0.439	0.435	0.401	0.397	0.027
394	0.027	0.027	45.568	45.445	0.331	0.329	0.355	0.352	0.005
403	0.051	0.051	69.476	69.254	0.304	0.302	0.356	0.353	0.009
430	0.045	0.045	73.02	73.052	0.375	0.373	0.372	0.369	0.009
438	0.075	0.075	69.088	69.134	0.437	0.435	0.401	0.399	0.069
440	0.069	0.067	90.171	89.89	0.416	0.414	0.391	0.388	0.028
462	0.057	0.057	88.844	88.903	0.398	0.396	0.381	0.379	0.012
475	0.035	0.035	66.428	66.109	0.339	0.335	0.355	0.352	0.003
479	0.078	0.078	96.561	96.137	0.427	0.422	0.398	0.395	0.042
498	0.047	0.047	63.264	62.915	0.324	0.32	0.357	0.353	0.015
502	0.033	0.031	52.135	51.938	0.267	0.265	0.338	0.334	0.005

We use consensus analysis to consider above parameters, import data from table 2 by Ucinet, and obtain the result (Answer Key) of consensus analysis as follows:

Table3. consensus analysis

rank	1	2	3	4	5	6	7	8	9	10
code	10	165	44	148	378	438	479	187	440	287

Thus it can be seen that. number 10 is the top 1. That is to say, ALON, NOGA M. has significant influence within the network.

Core. We can use UCINET and get the results as shown in Table 2:

Table 4. core

mark	name	core
10	ALON, NOGA M.	0.271
378	RODL, VOJTECH	0.241
44	BOLLOBAS, BELA	0.215
165	GRAHAM, RONALD LEWIS	0.209
148	FUREDI, ZOLTAN	0.206
479	TUZA, ZSOLT	0.19
177	GYARFAS, ANDRAS	0.179
440	SPENCER, JOEL HAROLD	0.178
462	SZEMEREDI, ENDRE	0.176
128	FAUDREE, RALPH JASPER, JR.	0.163

Just similar to the previous step, we choose the ranks of authors with top10 largest core values are shown in Table 2. Recommended core membership is top 40 nodes and concentration is 0.862. That is to say, these 40 authors occupy a core position and play an active role in the co-author networks.

Of course, we can get a ranked table of degree and the result will be similar to the first two tables. No matter what way we choose, there is only our measurable criteria are different. For example, we have to use betweenness when we examine the impact of intermediaries. As the result shows *HARARY, FRANK** is vital for the net. The fellow is the network just regard *HARARY, FRANK** as nodes central.

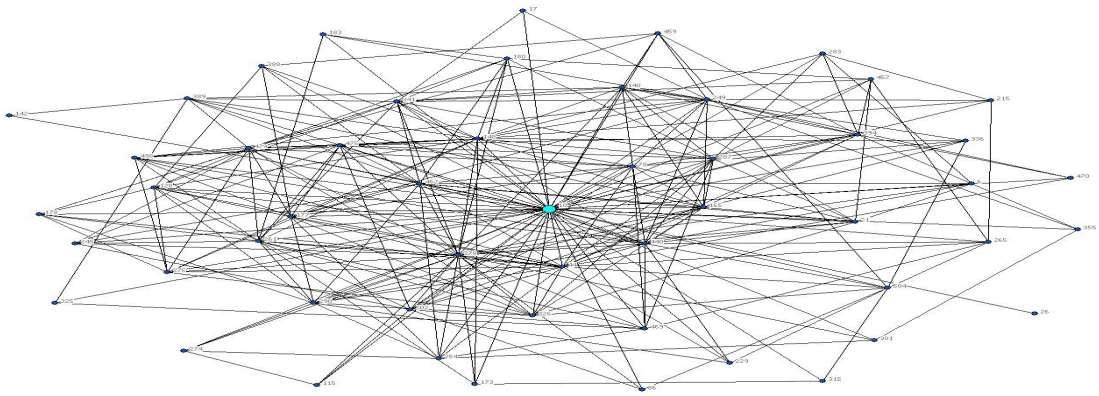

Figure 3. nodes central

Reference

[1] X. Liu, J. Bollen, M.L. Nelson, et al. Co-authorship networks in the digital library research community, Information Processing and Management 41 (2005) 1462-1480.

[2] H. Hou, H. Kretschmer, Z. Liu, The structure of scientific collaboration networks in Scientometrics, Scientometrics 75 (2008) 189-202.

[3] M.E.J. Newman, Scientific collaboration networks. I. Network construction and fundamental results. Physical Review E 64 (2001) 1-8.

[4] H. Kretschmer, Author productivity and geodesic distance in bibliographic co-authorship networks, and visibility on the Web. Scientometrics 60 (2004) 409-420.

[5] Matthew O. Jackson, Social and Economic Networks, Princeton University Press, 2008.

[6] J.D. Luo, Social Network Analysis (the second edition), Social Sciences Academic Press, 2003.

[7] P. Holme, S.M. Park, B.J. Kim, C.R. Edling, Korea university life in a network perspective: dynamics of a large affiliation network, Physica A 373 (2007) 821-830.