

Apply Social Network Analysis and Co-author Network in Ranking Interdisciplinary Authors

Yi-Ning Tu^{1*} and Chi-Hao Lin²

Department of Statistics and Information Science, Fu Jen Catholic University, Taiwan
eniddu@mail.gmail.com¹, howard81527@gmail.com²

*Corresponding Author

Received 25 February 2019; received in revised form 1 April 2019; accepted 13 May 2019

Abstract

This study tries to exploit the betweenness centrality of social network analysis and co-author network to rank the interdisciplinary authors. The interdisciplinary author is defined as an author who published papers at least in more than one research domain. This study discovers the interdisciplinary authors should have higher betweenness centrality than others. The interdisciplinary topics will apply in the "text mining" and "library and information science" that are the techniques used in this work. Finally, this study plan to compared the Spearman's rank correlation coefficient between h-index, total citation frequency and betweenness centrality with the proposed method.

Keywords: Betweenness centrality, Ranking interdisciplinary author, Social network analysis, Co-authors Network

1. Introduction

There are lots of research field are not only include one domain currently. Jacobs (2009) claims there are lots of successful research are interdisciplinary work. Take statistics as an example, no matter in biology (Liu, Pouli, Sood & Sundarakrishnan, 2017) or finance (Lorenz & Sachs 2016) there are some interdisciplinary and excellent studies. While there are some scholars want to investigate a new or combine another field as interdisciplinary, they need to spend plenty of time and effort to study lots of literature. Some studies proposed methods to help the researchers who want to combine another new field in their works. Burt (1980) suggests although the quality of papers and research domains are not easily evaluated, there are still some studies try to help this problem and proposed some indices for researchers. Such as finding out impact authors (Börner, Maru & Goldstone, 2004, discovering impact papers (Liu & Lu, 2012) or searching from impact journals (Garfield, 2006).

In impact authors, during last decade, the famous H-index (Hirsch, 2005) is a frequently used method for evaluating an author's impact. For evaluating journals, there is well-known Impact Factor(IF) proposed by Garfield (1955). As mentioned to papers or topics, Tu & Seng (2012) also assert that impact papers can be evaluated by impact powers of an author and their social connections. After that, they also proposed some measurement no matter to detect emerging topic or finding impact (Tu &

Seng, 2012). Also, they exploit the aging theory to develop some indices to help find the emerging topic (Tu & Seng, 2012). In 2016, (Tu & Hsu, 2016) proposed a conceptual trajectory development map, which can propose a fast browse method to trace the development of a research field by time. They start to trace the development of a research topic to help researchers find the possibility of interdisciplinary topics by the conceptual map. But these related works are focused on one topic. And there are no too much works discuss ranking the interdisciplinary authors.

Aboelela (2007) indicates although there are more and more interdisciplinary works than before, since the difference of the research domains, there is no effective approach to identify what is interdisciplinary paper. Porter (2009) exploits the categories defined by the Web of Science search engine as the different topic. This study adopts the same method to classify the category of a research field. Suppose D_h as a the h^{rd} domain. If there is an author published paper in D_1 or co-author with other people, besides, also have the same situation in D_2 . This study defines this author is an interdisciplinary author.

As mentioned to the impact of a paper or research topic, it's limited by time. The earlier published work will have higher chance to get higher citations. Hence, sometimes, when a topic starting to be discussed everywhere, it may over obsolete and not potential emerging topic anymore. This study exploits use the impact author as a starting point, especially the interdisciplinary papers may have chance be

the emerging topics. Consequently, this work tries to ranking the interdisciplinary authors to help the new comer to get into a research field easily and quickly. The new researchers can trace the trajectory and have chance discover a potential impact topic of a field by following the work of high ranking interdisciplinary author.

Diallo, Lynch, Gore & Padilla (2016) suggested that the importance of a paper although related to time, but if using centrality of social network analysis (SNA) to calculate will prevent the time effect. There are lots methods to compute the centrality of an author. But betweenness centrality can reflect the interdisciplinary authors more than other normal authors. When two people transmit the information, knowledge or extended the trajectory of a research, it's not always directly, sometimes it needs some literatures connected networks to passing the information or knowledge. The higher betweenness centrality author will have higher chance to be passed by his research than others when some studies want to transmit information or knowledge. This work tries to evaluate disciplinary authors and ranking these authors by their co-authorship network.

2. Literature Review

2.1 Co-author Network

Co-author network is a social network extension and has been broadly applied in understanding the structure of science field and author position. Co-author network and citation networks have similar principle. Both are the application in social network of different data. The differences are the citation network is using the paper as a node to express the cited and citing relationships while the co-author network represents nodes as authors. Except the citation relationship between papers, there are also different sort of networks. Reffay & Chanier (2003) suggest can construct the network between students groups and help the learning efficiency. Park (2003) asserts that social network can be applied in the hyperlink relationships between webs. Liu, Bollen & Nelson (2005) point out the citations between papers can be applied in citing and cited relationship. Besides, co-author network is more emphasize the cooperation in a paper between authors. Consequently, the co-author network is more similar social network than citation network.

Co-author network is based on an author as a node, and investigate the relationship between authors. Constructing the co-author network is expressed as formula (1). If there are co-author relationship between A_i and A_x , the $E_{i,x}$ will be 1, otherwise 0. Finally establish the co-author network G for every author in a field.

$$G=(A_i, E_{i,x}) \quad (1)$$

- $G_{i,x}$: If A_i and A_x exist the co-author relationship then build the connection
- A_i : the i^{th} author
- A_x : the other authors except A_i
- $E_{i,x}$: the co-author relationship between A_i and A_x

This study will download the papers from Web of Science and retrieve all the authors in each paper. Based on the formula (1) to drown the co-author network and compute the betweenness centrality of each author.

2.2 Social Network Analysis(SNA)

The social network defined as some nodes to connect as a network based on their relationships. These nodes could be personal or social groups, and the relationships can also apply in other field. For example, in the travel industry (Casanueva, 2016), patent (Sternitzke, 2008), and family relationship (Crosbieburnett, 1984), even to look for impact authors in each field (Otte & Rousseau, 2002) are all apply in social network analysis. Crucitti(2006) suggests that social network analysis(SNA) are include six different centralities such as betweenness, closeness, degree, efficiency, straightness, information. The SNA treats all the relationship as a link and analyzes the relationship. While a node in the network has more excellence than other position, it can also have more resources than others. Hence, the position and the connection will be a point of a network. Freeman (1978) proposed three import centrality which can measure the importance of the node in the network. They are degree centrality, betweenness centrality, closeness centrality. Yin, Kretschmer & Hanneman(2006) use these three centralities to compute the co-author network of COLLNET, and determine which paper is more important. The following this study give an example as table 1 to introduce the betweenness which is most suitable for computing the rank of interdisciplinary authors.

Suppose the A_i represent the i^{th} author, P_j indicates the j^{th} paper, and the table 1 shows P_j is published by authors A_1, A_3 , in other words, A_1, A_3 are co-authors. Viewing the node as authors and drawing the graph of relationship as figure 1 for table 1.

Table 1: The Relationship between Papers and Authors

Papers (P_j)	Authors (A_i)
P_1	A_1, A_3
P_2	A_3, A_4, A_5
P_3	A_1, A_2
P_4	A_4, A_5, A_6

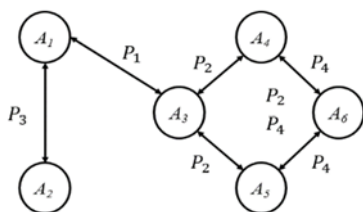


Figure 1: Co-author Network

2.3 Betweenness Centrality

Betweenness centrality represents as two nodes need a node to play a role as a bridge. While a node in the network has more excellence than other position, it can also have more resources than others. Hence, the position and the connection will be a point of a network. As the formula (2), A_y indicates an author except A_i and A_x . When A_x wants to transmit the information A_y , and the shortest path has to pass the node A_i , then the A_i will be the bridge author between A_x and A_y .

$$b_{xy}(A_i) = \frac{g_{xy}(A_i)}{g_{xy}} \tag{2}$$

- A_y : indicates an author except A_i and A_x .
- g_{xy} : represents how many shortest path between A_x and A_y .
- $g_{xy}(A_i)$: represents how many shortest path between A_x and A_y and through by A_i .

Like formula (3), the A_i will compute to with each pair of A_x and A_y as $b_{xy}(A_i)$. After summarizing all the $b_{xy}(A_i)$ will produce the $C_B(A_i)$. Using the same approach for each authors can compute the value of each author's betweenness centrality.

$$C_B(A_i) = \sum_{x=1}^{I-1} \sum_{y=1}^{I-1} b_{xy}(A_i) \text{ where } x < y \tag{3}$$

- $b_{xy}(A_i)$: the probability which A_i appear in the path between A_x to A_y .
- I : the number of all the authors.
- $C_B(A_i)$: the betweenness centrality of A_i .

Take figure 2 as an example, while A_1 want to transmit the information to A_5 has to pass the node A_3 . Hence A_3 has more power between A_1, A_3 and A_5 . In other words, while A_3 does not appear in the shortest path between A_1 and A_2 and the value of betweenness centrality of A_3 for path A_1 to A_2 will be 0. The frequency of A_3 in each path will show in the table 2.

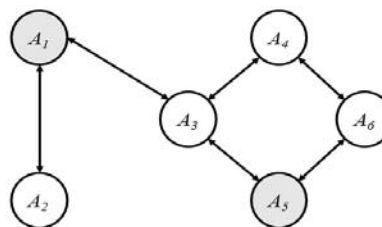


Figure 2: Computing the Betweenness Centrality

Table 2: The Frequency of A3 Appear in the Shortest Path

path	Probability in path	path	Probability in path
A_1 to A_2	0/1=0	A_2 to A_1	0/1=0
A_1 to A_4	1/1=1	A_4 to A_1	1/1=1
A_1 to A_5	1/1=1	A_5 to A_1	1/1=1
A_1 to A_6	2/2=1	A_6 to A_1	2/2=1
A_2 to A_4	1/1=1	A_4 to A_2	1/1=1
A_2 to A_5	1/1=1	A_5 to A_2	1/1=1
A_2 to A_6	2/2=1	A_6 to A_2	2/2=1
A_4 to A_5	0/1=0	A_5 to A_4	0/1=0
A_4 to A_6	0/1=0	A_6 to A_4	0/1=0
A_5 to A_6	0/1=0	A_6 to A_5	0/1=0

The table 3 shows the betweenness centrality for all the nodes of the co-author networks in the figure 2. Finally, this study will use the approach mentioned above to rank the interdisciplinary authors.

Table 3: The Betweenness Centrality for All the Nodes

A_i	$C_B(A_i)$	Rank
A_1	8	2
A_2	0	4
A_3	12	1
A_4	3	3
A_5	3	3
A_6	0	4

3. Interdisciplinary Authors Ranking Algorithm

This study tries to divide the process into three parts: 3.1 Construct the co-authorship network. 3.2 Compute the betweenness centrality of interdisciplinary authors. 3.3 Raking the interdisciplinary authors. The details are illustrated as follows.

3.1 Construct the Co-Authorship Network

3.1.1 Retrieve the Papers from Web of Science

In the academic paper search engines, Web of Science to search two research domains. Because it's interdisciplinary author ranking, hence the two domains can't too different to

find the co-authors. This study plans to use domains such as “text mining” and “library and information science” they are not only have the cross over parts and are also the techniques used in this study.

3.1.2 Extract the Information from Authors and Papers

Suppose there are two domains such as D_x and D_y . After retrieving all the papers in this two domains and present as P . Suppose there are 10 papers in these two domains shows as table 4. Since the P_5 is the paper across the D_x and D_y , hence the A_5 will be viewed as interdisciplinary author. Besides, the A_1 and A_4 are co-authors both in D_x and D_y are also viewed as interdisciplinary authors. Finally the A_1, A_4, A_5 are the interdisciplinary authors in the case.

Table 4: The Relationship between Domains, Papers and Authors.

D_h	P_j	A_i
D_x	P_1	A_1, A_3, A_4
D_y	P_2	A_1, A_2, A_4
D_y	P_3	A_4, A_5
D_x	P_4	A_5, A_6
D_x, D_y	P_5	A_5
D_y	P_6	A_7, A_8
D_y	P_7	A_8, A_9
D_x	P_8	A_{10}, A_{11}
D_y	P_9	A_{12}
D_x	P_{10}	A_{13}

3.1.3 Build the co-authorship network

Use the 10 papers and the 13 authors in table 4 to build the co-author network as the fig. 3. For example, A_1, A_3, A_4 these three authors are co-author in P_1 and can have link with each other. There are three major co-author networks. $A_1, A_2, A_3, A_4, A_5, A_6$ is the biggest one, and A_{12}, A_{13} without any co-authors and also not the interdisciplinary authors in these papers.

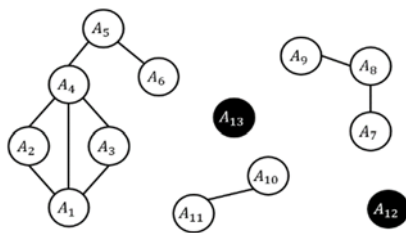


Figure 4: Building the Co-author Networks

3.1.4 Filter the Max Co-Authorship Network

Next step is filtering the max co-authorship network. This study labels the three networks as fig. 5. Besides, this study also computes the size of each network and show the ratio as table 5. Finally, this work will choose the largest network to finding out the interdisciplinary authors.

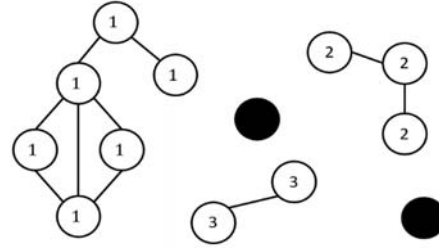


Figure 5: Label the Co-author Networks

Table 5: The Ratio of Each Co-author Network

No.	quantity	Occupied ratio
1	5	6/13=46.2%
2	3	3/13=23.1%
3	2	2/13=15.4%

3.1.5 Find out the Interdisciplinary Authors

After filtering the largest co-author network as network 1, all the authors in network 1 are listed as table 6. There are three authors such as A_1, A_4, A_5 are across two domains and will be viewed as interdisciplinary authors (IA_g) in this work.

Table 6: Find Out the Interdisciplinary Authors

A_i	D	IA_g
A_1	D_x, D_y	IA_1
A_2	D_y	-
A_3	D_x	-
A_4	D_x, D_y	IA_2
A_5	D_x, D_y	IA_3
A_6	D_x	-

3.2 Compute the Betweenness Centrality of Interdisciplinary Authors

The calculation method has introduced as section 2.3. The table 7 shows the betweenness centrality of author A_1 .

Table 7: Shows the Betweenness Centrality of Author A_1 .

Path	Probability in path	Path	Probability in path
A_2 to A_3	$1/2=0.5$	A_3 to A_2	$1/2=0.5$
A_2 to A_4	$0/1=0$	A_4 to A_2	$0/1=0$
A_2 to A_5	$0/1=0$	A_5 to A_2	$0/1=0$
A_2 to A_6	$0/1=0$	A_6 to A_2	$0/1=0$
A_3 to A_4	$0/1=0$	A_4 to A_3	$0/1=0$
A_3 to A_5	$0/1=0$	A_5 to A_3	$0/1=0$
A_3 to A_6	$0/1=0$	A_6 to A_3	$0/1=0$
A_4 to A_5	$0/1=0$	A_5 to A_4	$0/1=0$
A_4 to A_6	$0/1=0$	A_6 to A_4	$0/1=0$
A_5 to A_6	$0/1=0$	A_6 to A_5	$0/1=0$

3.3 Ranking the Interdisciplinary Authors

The betweenness centralities of all the authors are shown as table 8. The computing method has shown in section 2.3.

Table 8: The Betweenness Centrality of All the Authors

A _i	C _B (A _i)	Rank
A ₁	1	3
A ₂	0	4
A ₃	0	4
A ₄	13	1
A ₅	8	2
A ₆	0	4

4. Experimental Results

This research selects the two domain which is “text mining” and “Library and In-

formation Science”. The data was collected during 2000/1/1~2017/1 and the detail are shown as table 9. The largest co-author network includes 3346 authors, and the second largest contains 135 authors. Hence the largest co-author network is more impact. The top 10 between centrality of authors from the selected network are shown in the table 10. Besides, the network contains 60 interdisciplinary authors and shown as table 11.

And contains 60 authors are interdisciplinary authors.

Table 9: The Detail of the Data Collection

Domain(D)	Text Mining	Library and Information Science
Total paper quantity	4445	3752
Total author quantity	13573	10778
The average author numbers in each paper	3.05	2.87

Table 10. The Betweenness Centrality of Top 10 Authors in the Largest Network

Author’s name	Betweenness centrality	interdisciplinary authors
Verspoor, Karin	2165618	-
Cohen, Kevin Bretonnel	1851809	-
Jegga, Anil G.	1809885	-
Wang, Lishan	1777581	-
Song, Min	1675064	O
Xing, Li	1450623	-
Chen, Xi	1450542	-
Yan, Erjia	1393681	O
Liu, Yufeng	1341488	-
Li, Xia	1260355	-

Table 11: The Rank of Betweenness Centrality for All Interdisciplinary Authors

Author’s name	Betweenness centrality
Song, Min	1675064
Yan, Erjia	1393681
Xu, Hua	737280
Wu, Fei	660960
Chen, Kun	616930
Xu, Tao	591682
Zuccala, Alesia	541346.7
Zhang, Yin	528282
Ding, Ying	525459.1
Yu, Philip S.	460344
Liu, Yang	309730
Buttigieg, Pier Luigi	279836.5
Zhang, Lei	264684
Li, Jiang	259171.9
Kalas, Matus	205282
Tang, Jie	179034
Luo, Hui	126274
Evans, James A.	119698
Li, Gang	117838.2
Laibe, Camille	113050
Rzhetsky, Andrey	106528
Pavlopoulos, Georgios A.	97233.92
Antezana, Erick	93184
Wu, Lei	86554
Malone, James	73282
Li, Jun	66690
Wang, Ping	66660
Li, Lang	59994
Schulz, Stefan	59994
Thelwall, Mike	53378

Author’s name	Betweenness centrality
Suzek, Baris E.	53344
Janssens, Frizo	46732
Valentin, Franck	46690
He, Wu	40048
Kuentzer, Jan	40032
Kovacevic, Aleksandar	26715.33
Glanzel, Wolfgang	26704
Mu, Xiangming	26704
Zhang, Yan	26704
Zhang, Ying	26704
Cronin, Blaise	26704
Peitsch, Manuel C.	26704
Li, Jiang	20034
Iliopoulos, Ioannis	17221.22
Kell, Douglas B.	13360
Luo, Bo	13360
Wang, Xue	13360
Wang, Yuan	13360
Zhang, Jing	13360
Hoffman-Apitius, Martin	8276.51
Bai, Xue	6682
Kohane, Isaac	6682
Chen, Wei	6682
Williams, Antony J.	6682
Younesi, Erfan	85.32381
Heo, Go Eun	7.666667
Liu, Xiaomin	0
Wu, Lili	0
Huvila, Isto	0
Zhu, Yongjun	0

5. Conclusion

This research proposed a new method to solve the problem of ranking interdisciplinary authors. The proposed method combining the betweenness centrality of social analysis and trying to construct the co-author networks to rank the interdisciplinary authors. There are 60 interdisciplinary authors were rank from the domain “text mining” and “library and information science”. The top author in these two domains is Song, Min whose h index is rank as 9 in the total authors and the total citations are rank 16. These information tell the original method to evaluate the authors are similar to the proposed method. But the proposed method can really reflect the interdisciplinary author’s rank.

References

- Aboelela, S. W., Larson, E., Bakken, S., Carrasquillo, O., Formicola, A., Glied, S. A., & Gebbie, K. M. (2007). Defining interdisciplinary research: Conclusions from a critical review of the literature. *Health Services Research, 42*(1), 329-346.
- Börner, K., Maru, J. T., & Goldstone, R. L. (2004). The simultaneous evolution of author and paper networks. *Proceedings of the National Academy of Sciences, 101*, 5266-5273.
- Burt, R. S. (1980). Autonomy in a social topology. *American Journal of Sociology, 85*(4), 892-925.
- Casanueva, C., Gallego, Á., & García-Sánchez, M. R. (2016). Social network analysis in tourism. *Current Issues in Tourism, 19*(12), 1190-1209.
- Crosbie-Burnett, M. (1984). The centrality of the step relationship: A challenge to family theory and practice. *Family Relations, 33*, 459-463.
- Crucitti, P., Latora, V., & Porta, S. (2006). Centrality measures in spatial networks of urban streets. *Physical Review E, 73*(3), 036125.
- Diallo, S. Y., Lynch, C. J., Gore, R., & Padilla, J. J. (2016). Identifying key papers within a journal via network centrality measures. *Scientometrics, 107*(3), 1005-1020.
- Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks, 1*(3), 215-239.
- Garfield, E. (1955). Citation indexes for science. *Science, 122*, 108-111.
- Garfield, E. (2006). The history and meaning of the journal impact factor. *Jama, 295*(1), 90-93.
- Hirsch, J. E. (2005). 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), 16569-16572.
- Jacobs, J. A., & Frickel, S. (2009). Interdisciplinarity: A critical assessment. *Annual Review of Sociology, 35*, 43-65.
- Liu, J. S., & Lu, Y. Y. L. (2012). An integrated approach for main path analysis: development of the Hirsch index as an example. *Journal of the American Society for Information Science and Technology, 63*(3), 528-542.
- Liu, X., Bollen, J., Nelson, M. L., & Van de Sompel, H. (2005). Co-authorship networks in the digital library research community. *Information Processing & Management, 41*(6), 1462-1480.
- Liu, Z., Pouli, D., Sood, D., Sundarakrishnan, A., Mingalone, C. K. H., Arendt, L. M., & Schnellendorfer, T. (2017). Automated quantification of fiber-like structures in biological tissues. *Biomaterials, 116*, 34-47.
- Lorenz, N., & Sachs, D. (2016). Identifying laffer bounds: a sufficient-statistics approach with an application to Germany. *The Scandinavian Journal of Economics, 118*(4), 646-665.
- Otte, E., & Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science, 28*(6), 441-453.
- Park, H. W. (2003). Hyperlink network analysis: A new method for the study of social structure on the web. *Connections, 25*(1), 49-61.
- Porter, A., & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics, 81*(3), 719-745.
- Reffay, C., & Chanier, T. (2003). How social network analysis can help to measure cohesion in collaborative distance-learning. In *Designing for Change in Networked Learning Environments*, 343-352. Springer Netherlands.
- Sternitzke, C., Bartkowski, A., & Schramm, R. (2008). Visualizing patent statistics by means of social network analysis tools. *World Patent Information, 30*(2), 115-131.
- Tu, Y. N. & Seng, J. L. (2012). Emerging topics detection and measurement. *Advanced Science Letters, 9*, 653-659.
- Tu, Y. N. & Seng, J. L. (2012). Indices of Novelty for Emerging-topic Detection. *Information Processing and Management, 48*, 303-325.
- Tu, Y. N., & Hsu, S. L. (2016). Constructing conceptual trajectory maps to trace the

development of research fields. *Journal of the Association for Information Science and Technology*, 67(8), 2107-2118.

Tu, Y. N., & Seng, J. L. (2011). Evaluating the impact power of authors via Bayesian estimation of authors' social connections. *The 2011 International conference on Advances in Social Networks Analysis and Mining*, 678-684.

About Authors

Yi-Ning Tu is an associate professor whose Ph.D. degree in Management of Information System. Her research interests include data mining, text mining and decision support systems for big data-related problems.

Chi-Hao Lin is graduated from the master degree of the Department of Statistics and Information Science. He is a student of the professor Tu. His research interests are social network analysis and co-author ranking in interdisciplinary.