

---

# Exploring Homophily in Research Collaboration: A Dynamic Centrality Analysis Approach

---

Laxminarayan Sahoo\* and Sanchita Guchhait

*Department of Computer and Information Science; Raiganj University;*

*Raiganj-733134, India*

*E-mail: lxsahoo@gmail.com; mou.guchhait@gmail.com*

*\*Corresponding Author*

Received 12 May 2024; Accepted 28 July 2024

## **Abstract**

Homophily is the phenomenon of individuals seeking others who are similar to themselves. Homophily influences the formation of co-authorship networks. In our study, we measure the homophily of the authors based on their affiliation using the co-authorship network. The main contribution of our study is that we test homophily with a dynamic centrality analysis algorithm and find that homophily exists when we measure the authors' degree within and outside their network. However, homophily does not exist when we use the dynamic centrality analysis algorithm for the same co-authorship network.

**Keywords:** Co-authorship network, dynamic centrality, homophily, social network.

*Journal of Graphic Era University, Vol. 12.2, 243–262.*

doi: 10.13052/jgeu0975-1416.1224

© 2024 River Publishers

## 1 Introduction

Social networks are ubiquitous and essential in our daily lives. They also offer opportunities for researchers to share their work and connect with other professionals. ResearchGate, for example, is a social network for scientists to exchange scientific knowledge and stay updated on the latest developments in their field. A social network consists of nodes that are connected by some type of relationship, such as common interests, affiliations, or collaborations [1]. One of the key factors that influence the formation of relationships in a social network is homophily. Homophily is the tendency of similar individuals to associate and interact more than dissimilar ones [2]. Homophily may be based on various attributes, such as age, gender, education, social class, language, colleagues, occupation, religion, interest, etc. [3]. In the academic domain, homophily also affects the creation of co-author ties, which may depend on factors like affiliation, gender, institution, experience, country, etc. Homophily has an important implication for analyzing online communities and understanding various social phenomena, such as perception biases, segregation, inequality, and information diffusion among different groups of individuals [4]. Recently, homophily has also been incorporated into a neural network model that combines network and textual features to improve link prediction and occupation prediction in social networks [3]. In the co-authorship network, the centrality of each author reflects their influence and status in the network. Centrality measures the number or quality of connections that a node has within a network. Identifying the most central nodes in a social network has many applications, such as finding the most influential individuals, disseminating information, preventing disease spread, etc. [5, 17]. Homophily is the tendency of individuals to link with others who have similar thinking, characteristics, or experiences. In the field of research collaboration, coauthors are also connected with other individuals who have the same field of research area or the same beliefs/affiliations. There are some important implications for academic diversity and innovation from this phenomenon. Research collaborators may feel comfortable talking and understanding one another when they have similar educational backgrounds, work experiences, or cultural perspectives. This simplicity of communication may promote a more pleasant working environment where miscommunication is reduced and ideas are allowed to flow more freely. As the team is homophilic therefore the outcome of research work may be more productive. Because the coworkers can collaborate more effectively and peacefully.

Furthermore, the decision-making process may be simplified easily by team members. Collaborators are more likely to agree on the course of their research and the appropriate methodologies when they share the same presumptions, values, and objectives. In a co-authorship network as the team members are homophilic, they may spend less time debating and negotiating and more time working on the project at hand. As there does not any miscommunication or arguments among the collaborators, therefore, the project completion speed and efficiency will be improved. These advantages might be more helpful in research settings when effectiveness and time management are crucial.

Many studies have investigated on homophily and centrality in co-authorship networks using different methods and metrics, such as citation counts, h-indexes, etc. However, the impact of homophily on centrality has not been well explored. In this paper, we examine how homophily affects dynamic centrality for each author in the co-authorship network. The main contributions of this paper are as follows:

- We review the concept and role of homophily in social networks.
- We propose a method for detecting homophily in a social network.
- We introduce dynamic centrality and analyze how it relates to homophily in co-authorship networks.
- We compare the results of normal homophily and dynamic centrality-based homophily.

The rest of the paper is organized as follows. Section 2 presents the related work in this domain. Sections 3 and 4 introduce some preliminaries of homophily and types of centrality measures respectively. Section 5 describes the dataset used for our experiments. Sections 6 and 7 present the algorithms for homophily detection and dynamic centrality analysis, respectively. Section 8 discusses the results and findings of our study. Section 9 concludes the paper and outlines some limitations and future directions.

## **2 Literature Review**

Homophily is the phenomenon of interacting with similar individuals, which has been studied in various socio-psychological contexts, such as community development, segregation, social mobility, etc. [1, 6] investigated how homophily and network structure affect the citation counts of individual authors in co-authorship networks. They collected data from the sociology departments of three East-European countries: Romania, Poland, and

Slovenia. In these co-authorship networks, each sociologist or author was represented as a node or ego, while their co-authors were considered as alters. They performed descriptive statistics and hierarchical regression models for each country separately and found that co-authors' citations had a significant impact on the authors' citations. One of the limitations of this paper is that they only used quantitative methods, which may not be sufficient to obtain a better result. Kwiek et al. and Holman et al. [7, 18] showed that gender homophily in research collaboration is important. The gender homophily theory states that male researchers tend to collaborate with other males, but female researchers do not prefer to form ties with other females. They created a large database called "The Polish Science Observatory" by combining the Polish academic scientist database and Scopus database, which contained comprehensive, administrative, and integrated biographical information. They used a fractional logit regression approach to analyze the result and found that gender homophily was more prevalent for male researchers than female researchers. They also achieved 100% gender determination for all researchers in the system. Co-authorship networks are formed when two or more authors collaborate on one or more papers. To measure homophily in a co-authorship network, it is necessary to measure the centrality of each author in the network Lu et al. and Cristani et al. [5, 14]. Several researchers have proposed various centrality measures based on different aspects, such as reachability, current flow, feedback, vitality, random process, etc. Das et al. [8]. The most commonly used centrality measures are degree centrality, betweenness centrality, and closeness centrality. The degree centrality of a node measures the number of nodes connected to it. The betweenness centrality of a node measures the number of shortest paths that pass through it. The closeness centrality of a node measures the average distance from that node to all other nodes Das et al. and Bloch et al. [8, 9]. McPherson et al. and Lawrence et al. [10, 11] classified homophily into two types: value homophily and status homophily based on different attributes, such as gender, ethnicity, social status, age, etc. They also mentioned that geographical distance is also important for homophily. Dwivedi et al. and Gallivan et al. and Fagan et al. [12, 13, 19] constructed a co-authorship network where the nodes were information science researchers and the linking criterion was the co-authorship of an article. They used data from different eight journals on the business analytics discipline. To analyze the network, they used Exponential Random Graph Modeling [ERGM] and the results suggested that these researchers were more likely to collaborate with other similar scholars.

Bisgin et al. [2] experimented with whether it is possible to form new connections in social media based on the principle of homophily as in the real world. For their experiment, they studied three different online social media sites: Last.fm, BlogCatalog, and LiveJournal and introduced different methodologies: dyadic relation, random rewired, community-based, and content-based analysis. Their experimental result showed that interest-based homophily did not play an important role in creating new connections on social media sites.

Umadevi and Dias et al. [15, 16] focused on centrality measures in co-authorship networks in management and accounting science domains. Their experiment showed that centrality was crucial for identifying the influential nodes in a network. The literature provides valuable insights into the area of centrality analysis, with works such as R. Mahapatra et al. [21] introducing how to find influential nodes in a network using a neutrosophic graph, how much influences occur in the social network is described by S. Samanta et al. [22].

Jones et al. [20] proposed a concept of homophily and dynamic centrality in LGBQ (Lesbian, Gay, Bisexual, Questioning) students and found that LGBQ students did not experience the same social exclusion as previous generations. G. Muhiuddin et al. [23] focused on centrality measures of different diseases due to DNA sequencing. S. D. Pandey [24] et al. measure the centrality in bipolar-valued fuzzy graphs. The authors' contribution is given in Table 1.

### **3 Preliminaries**

The researchers classified the homophily into two types based on the relevance of homophily as McPherson et al. [10] - status homophily and value homophily.

#### **3.1 Status Homophily**

It means that the individuals with same social status are like to form a homophily with each other. It comprises both ascribed characteristics and acquired characteristics in society. Ascribed characteristic refers to the features of an individual acquired at birth or through inheritance typically sex, caste, height, ethnicity, etc. and the acquired characteristics may include behavior patterns, education, religion, occupation, etc. that are obtained after birth [10].

**Table 1** Development of work by different authors

Year	Authors	Contribution	Limitations
2012	H. Bisgin et al.	They introduced different methodologies like dyadic relation, random rewired, community based and content-based analysis	
2018	A.A. Stocia	How homophily and structural properties of the network influence the individual citation calculation in co-authorship networks.	Only quantitative method is proposed.
2020	R. Dwivedi et al.	Based on business analytics discipline, they build a co-authorship network which is analyzed using ERGM and the co-authorship links are predicted with the help of authors' departmental affiliations, continental affiliations etc.	
2021	M. Kwiek et. Al	<ul style="list-style-type: none"> <li>• They created a large database 'The Polish Science Observatory Database' which contains 99,535 scientists of the scopus database and 25,463 university professors.</li> <li>• They used a fractional logit regression approach for analyzing the result and found that gender homophily is more appropriate for male researchers than female researchers.</li> <li>• For all researchers in the system, achieved 100% gender determination.</li> </ul>	They select the authors who have the publication in the Scopus only for a certain period (2009–2018). The choice of different databases, various forms of publications, and different periods might result in other results.
Proposed work		Homophily in research collaboration with dynamic centrality analysis.	

### 3.2 Value Homophily

It considers the interaction with others who have similar thinking like attitudes, values, and beliefs, although there is a great difference between their social statuses [2].

## 4 Types of Centralities

Every network is considered as a graph, where the graph  $G(V, E)$  is a diagram consisting of a collection of vertices together with edges joining a certain pair of these vertices. Where  $V$  denotes the set of vertices and  $E$  denotes the set of edges. A graph is said to be connected if there exists a path between every pair of vertices. Otherwise, the graph is disconnected. Mathematically the relationship between two nodes  $m$  and  $n$  can be represented by

$$\begin{aligned} g_{mn} &= 1, \text{ if there exist a path between } m \text{ and } n \\ g_{mn} &= 0, \text{ otherwise} \end{aligned} \quad (1)$$

The geodesic distance  $d(m, n)$  from the node  $m$  to the node  $n$  is designed by  $d(m, n) =$  the count of links present in the shortest path from the node  $m$  to  $n$ , if there exists a path [8].

When in a graph or network, the node represents a group or a person and there exists a connection between them then this network is called a social network. To recognize the influential node in a social network centrality measure is very important. Network-based centrality measures can be classified on different aspects like reachability, shortest path, feedback, etc. Degree centrality, eccentricity centrality and closeness centrality, etc. come under the reachability-based centrality [8].

### 4.1 Degree Centrality

The number of edges that are directly connected to a particular node, denotes the degree centrality of that node. The degree centrality of a node is given by [9] –

$$C_D(i) = d_i \quad (2)$$

Where  $d_i$  is the degree of node  $i$ . To get a number from 0 to 1 its normalized form is

$$C'_D = \frac{d_i}{n-1}, \quad (3)$$

Where,  $n$  denotes the size of the network.

#### 4.2 Closeness Centrality

The network distance between a particular node and every other node is defined as the closeness centrality. The closeness centrality of a node  $i$  is defined by Das et al. [8]

$$C_c(i) = \frac{1}{\sum_{j \in N} d(i, j)} \quad (4)$$

Where  $N$  is the set of all vertices in the graph and  $d(i, j) =$  geodesic distance.

In normalized form, the closeness centrality is given by

$$C'_c(i) = \frac{n-1}{\sum_{j \in N} d(i, j)} \quad (5)$$

#### 4.3 Betweenness Centrality

In a network the betweenness centrality of a vertex  $i$  is defined by Das et al. [8]

$$C_B(i) = \sum_{x \neq y \in N} \frac{\partial_{xy}(i)}{\partial_{xy}} \quad (6)$$

Where,  $\sigma_{xy}(i) =$  count of the shortest path between  $x$  and  $y$  through vertex  $i$  and  $\sigma_{xy} =$  count of the whole shortest path between  $x$  and  $y$  in a network.

#### 4.4 Eigenvector Centrality

If the neighborhood of a node is significant then that node will be also significant. So, to measure the eigenvector centrality of a node depends on the entire neighborhood node's centrality. Eigenvector centrality of a vertex  $i$  is given by Bloch et al. [9].

$$Ax = \lambda x, \quad \lambda x_i = \sum_{j=1}^n a_{ij} x_j \quad (7)$$

Where  $i = 1, 2, 3, \dots, n$  and  $\lambda$  is constant and  $A$  is the adjacency matrix as  $a_{ij} = 1$ , if the vertex  $i$  is linked to the vertex  $j$  otherwise,  $a_{ij} = 0$ .



**Table 2** Representation of the dataset

University Name	Department	No. of Authors	No. of Co-authors	Total
Indian Institute of Technology Bombay	Computer Science and Engineering	42	10	420
Stanford University	Computer Science	66	10	660

## 5 Dataset

For experimenting with homophily, we have collected the data from DBLP (DataBase Systems and Logic Programming), which is a bibliographic database for computer science, and the period of collecting data from August to September 2022. Only the regular faculty of the computer science department of the two universities have been taken. The top 10 co-authors of every author have been taken with whom they publish their more articles. Out of these 10 co-authors, we compute the inter-network degree and the outside-network degree. Inter network degree is the number of co-authors who belong to the same university as the author, and the outside-network degree is the number of co-authors who are associated with different workplaces. The following Table 2. shows the complete representation of the dataset.

## 6 Homophily Detection Algorithm

In our experiment for finding the likelihood of association between the author and co-author, we developed a simple homophily detection algorithm, which is described below.

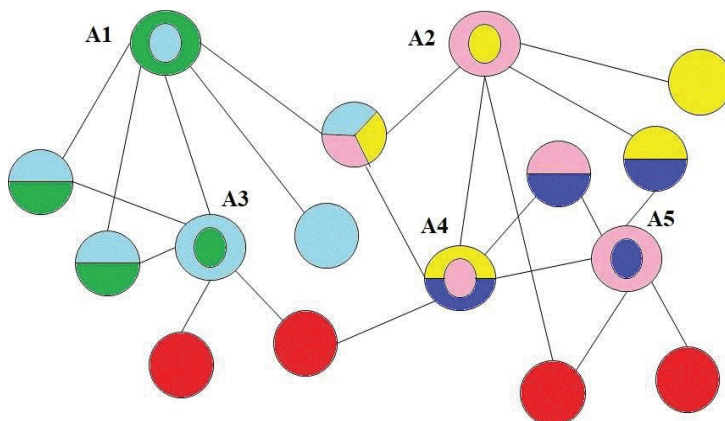
### 6.1 Algorithm

Input- Take a network.

Output- Does any homophily exist or not between the node and the particular network in which they belong to.

Steps:

- [1] For each node, select the top  $m$  collaborators and label them as either inter-network or outside-network, depending on whether they belong to the same network as the node or not.
- [2] Calculate the inter-network degree and outside-network degree for each node, which are the number of inter-network and outside-network collaborators respectively. Note that  $m$  is the sum of these two degrees.



**Figure 1** A co-authorship network.

**Table 3** Inter network and outside network degree of every author for Figure 1

Author	Inter Network Degree	Outside Network Degree
A1	5	0
A2	4	1
A3	3	2
A4	4	1
A5	3	2

- [3] Perform a t-test to compare the mean inter-network degree and mean outside-network degree across all nodes. The null hypothesis is that there is no difference between them.
- [4] If the p-value of the t-test is less than a significance level  $\alpha$  ( $\alpha = 0.05$  for this study), then reject the null hypothesis and conclude that there is homophily in the network. Otherwise, accept the null hypothesis and conclude that there is no homophily in the network.

Example 1: For the experiment, here we take five authors and five co-authors of each of them. The co-authorship network is shown in Figure 1. All authors are denoted by a center circle with different colors and all co-authors of respective authors are represented by the same color of that author. Here all co-authors are categorized by inter-network degree and outside-network degree. Every outside categorized co-author is denoted by the same color as red.

The inter and outside network degree for every author is given in Table 3, and their statistical analysis is also given in Table 4.

**Table 4** Statistical analysis with homophily detection algorithm of Figure 1  
t-Test: Paired Two Sample for Means

	Variable 1	Variable 2
Mean	3.8	1.2
Variance	0.7	0.7
Observations	5	5
Pearson Correlation		-1
Hypothesized Mean Difference		0
Df		4
t Stat	3.474396145	
P(T<=t) one-tail	0.012740741	
t Critical one-tail	2.131846786	
P(T<=t) two-tail	0.025481481	
t Critical two-tail	2.776445105	

According to our homophily detection algorithm, we find that the p-value is 0.012740741, which is less than  $\alpha$  ( $\alpha = 0.05$  for our study). Also, the mean of Variable 2 (outside network degree) is less than that of Variable 1 (inside network degree). Therefore, homophily exists for this network.

## 7 Homophily Detection With Dynamic Centrality Analysis Algorithm

The centrality measure plays a great role in identifying the influential or central node in a network. Basically, when the degree of a node is greater than all other nodes in a network, that node is identified as a central node. Homophily can be tested with the help of the centrality value, too. But here, we use the dynamic centrality value for testing the homophily.

Centrality in graph theory or network analysis refers to the measure of the importance or prominence of a node within a network. It quantifies how central or influential a node is based on its connectivity and position to other nodes. This centrality value may change depending on the different networks' characteristics. By the word dynamic, we understand changes of value depending on time. But in our study by the word dynamic centrality, we mean the changes of centrality value depending on different network features like connectivity, importance, or influence within the network.

For example, we consider two cities Delhi and Kashmir. We all know that Kashmir is a more beautiful place than Delhi. So, depending on different parameters like natural beauty, importance of the place, etc. Kashmir carries

a higher centrality value than Delhi. However, it is found that the greatest number of foreigners visit Delhi than Kashmir. That does not mean the centrality value of Kashmir is less. As the centrality value changes depending on different parameters of the network, therefore, it will be termed as dynamic centrality. But in that case, homophily does not exist. Again, if we let two places Kolkata and Delhi, then we find that the centrality value of Delhi is greater than Kolkata, and more visitors also visit Delhi and there exists homophily. The presence or absence of homophily, which refers to the tendency for nodes to connect to similar nodes, can further influence how centrality values are distributed within the network, adding another layer of complexity to understanding network dynamics.

While traditional centrality measures provide valuable insights into node importance within a static network, dynamic centrality extends this by capturing how centrality values change over in response to varying network conditions.

### 7.1 Algorithm

Input – Take a network.

Output – Is there any homophily that exists or not between the node and the particular network to which they belong to.

Step 1 – We are considering  $n$  nodes and top  $m$  collaborators of every node.

Step 2 – Then compute the inter-network degree and outside-network degree for every  $n$  node individually. Where the inter-network degree is the number of collaborators who are associated with the same network as authors, and outside network degree is the number of collaborators who are associated outside the network.

$$m = \text{inter network degree} + \text{outside network degree}$$

Step 3 – Measure the dynamic centrality for every  $n$  node inter-network degree and outside-network degree separately by

$$C = \frac{D(x) - Avg}{Max} \quad (8)$$

Where,  $D(x) = D_I(x)$  or  $D_O(x)$ , inter-network degree of a particular author or outside-network degree of a particular author.

$$Avg = Avg(I) \text{ or } Avg(O)$$

**Table 5** Inter network centrality and outside network centrality for the Figure 1

Author	Inter	Outside	Max (In)	Max (Out)	Total (In)	Total (out)	Avg (In)	Avg (out)	Centrality (In)	Centrality (out)
	Network Degree	Network Degree								
A1	5	0	5	2	19	6	3.8	1.2	0.76	-0.6
A2	4	1							0.04	-0.1
A3	3	2							-0.16	0.4
A4	4	1							0.04	-0.1
A5	3	2							-0.16	0.4

$$Avg(I) = \frac{\text{Total inter network degree}}{\text{No. of authors}} \quad \text{and}$$

$$Avg(O) = \frac{\text{Total outside network degree}}{\text{No. of authors}}$$

$$Max = Max(I) \text{ or } Max(O)$$

$Max(I)$  = Maximum value of an inter-network degree from all collaborators.

$Max(O)$  = Maximum value of outside network degree from all collaborators.

Step 4 – Perform statistical analysis to find the significant difference.

Step 5 – If the p-value is less than  $\alpha$  ( $\alpha = 0.05$  for this study) and the mean of Variable 2 (outside network degree) is less than that of Variable 1 (inside network degree) then the data contains homophily. Otherwise, there does not exist homophily.

**Example 2:** To test the homophily with dynamic centrality we also use the above network. The centrality of the inter-network degree and the centrality of the outside network degree for every author is computed and given in the following Table 5 and their statistical analysis is also given in the following Table 6.

According to our homophily detection algorithm with dynamic centrality we find that P-value is 0.390830468 which is greater than  $\alpha$  ( $\alpha = 0.05$  for our study). Therefore, when we test the homophily with dynamic centrality algorithm we find that there does not exist homophily for the above network.

## 8 Result Analysis

Generally, when the inter network degree for any node is higher than the outside network degree we can say that there exists a homophily between the node and the particular network in which the node is associated. This

**Table 6** Statistical analysis of Figure 1 with dynamic centrality analysis algorithm for detecting homophily

t-Test: Paired Two Sample for Means		
	Variable 1	Variable 2
Mean	0.104	0
Variance	0.14448	0.175
Observations	5	5
Pearson Correlation	-0.930762387	
Hypothesized Mean Difference	0	
df	4	
t Stat	0.296422639	
P(T<=t) one-tail	0.390830468	
t Critical one-tail	2.131846786	
P(T<=t) two-tail	0.781660937	
t Critical two-tail	2.776445105	

**Table 7** Statistical result of our experimental dataset without dynamic centrality analysis algorithm

t-Test: Paired Two Sample for Means		
	Variable 1	Variable 2
Mean	3.638889	6.361111
Variance	5.410436	5.410436
Observations	108	108
Pearson Correlation	-1	
Hypothesized Mean Difference	0	
df	107	
t Stat	-6.0812	
P(T<=t) one-tail	9.38E-09	
t Critical one-tail	1.659219	
P(T<=t) two-tail	1.88E-08	
t Critical two-tail	1.982383	

hypothesis is tested in our experiment with the help of a Two-sample t-test on the collected dataset. And the experimental result from the statistical analysis shows that the p-value is significantly less than the  $\alpha$ -value which is considered as 0.05 for our study and concludes that there exists homophily for that dataset. The mean of two sample variables, p-value, etc. are shown in the following Table 7.

But for the same dataset when we perform the t-test after applying the dynamic centrality analysis algorithm we observe that the p-value is highly

**Table 8** Statistical result of our experimental dataset with dynamic centrality analysis algorithm

t-Test: Paired Two Sample for Means		
	Variable 1	Variable 2
Mean	-1.60623E-18	-0.000815977
Variance	0.054104361	0.053785694
Observations	108	108
Pearson Correlation	-0.98271965	
Hypothesized Mean Difference	0	
df	107	
t Stat	0.018334496	
P(T<=t) one-tail	0.492703077	
t Critical one-tail	1.659219312	
P(T<=t) two-tail	0.985406153	
t Critical two-tail	1.98238337	

greater than the  $\alpha$ -value (0.05) and falls under the null hypothesis. Here for every node, both inter-network degree and outside-network degree the centrality is computed and that value is taken as the sample for the t-test. The following Table 8 shows the statistical result of two paired t-tests.

In our experiment, we collect the data from two research universities. One university belongs to India and another is outside India and we observed that in foreign universities, the maximum authors' outside network degree is higher than the internetwork degree because the maximum number of Indian researchers tends to do research work with foreign universities. That means it is not that the Indian research university is less valuable than the foreign university. Therefore, we compute the dynamic centrality for this type of network. Dynamic centrality is important for homophily analysis but does not apply to every network. Hence in our experimental network homophily exists when we compute the simple statistical analysis, but with dynamic centrality analysis, homophily does not exist.

In research collaboration, homophily may lead to more productive work and communication, but it also limits the range of ideas and perspectives that are essential to the diversity and uniqueness of the research. Using dynamic centrality analysis to study homophily in research networks is not without limitations. These could be the selections made for network attributes (such as connectedness and importance), which could introduce biases in the process of calculating dynamic centrality. If these standards are not properly chosen, the analysis may not accurately represent the true importance or centrality

of nodes within the network. Comparing dynamic centrality values across research networks may be difficult due to variations in the factors chosen and their relative importance. This can lead to biases when attempting to draw general conclusions from specific case studies or examples. Assessing centrality based on different features can be subjective, leading to biases. Different researchers might prioritize different features, resulting in varying centrality values that are not directly comparable. Besides that, dynamic centrality does not apply to all types of networks.

## 9 Conclusion

Homophily is an important concept in the construction of co-authorship networks, as it has applications in both the real world and social networks. Homophily studies have been conducted extensively in the fields of sociology, marketing, politics, etc. To test the presence of homophily in a co-authorship network, we used two algorithms. One is a homophily detection algorithm, and the other is a homophily detection algorithm with dynamic centrality analysis. The results from the simple homophily detection algorithm show that researchers tend to collaborate with other researchers who belong to the same affiliation. However, the results from the dynamic centrality-based homophily detection algorithm show that there is no homophily in the co-authorship network.

Some limitations of our study are that we only included the top 10 co-authors of each author who are regular faculty members in a single department at two universities. We selected the top 10 co-authors based on the number of publications they co-authored with each author. The results may change if we consider the entire co-author list for each author. Secondly, the dynamic centrality algorithm is useful for testing homophily but may not apply to all kinds of networks. Moreover, we only tested our hypothesis using the t-test. Other types of models can be developed to test this kind of hypothesis.

In the future, we plan to extend our study to different domains, such as business analytics, sports, etc. We also aim to investigate the existence of homophily in online social media networks, such as Twitter, Instagram, BlogCatalog, LiveJournal, etc.

Future studies can be done in some of the following areas to learn more about how homophily affects research collaboration: It can be examined how the geographical and cultural attributes affect in homophilic research collaboration network. It may determine the key nodes and the influential



researchers and then look at how homophily affects their role and relationship within the network. It can be analyzed how initial homophilous contacts change and impact the productivity and effectiveness of long-term collaborations. To acquire a qualitative understanding of the function of homophily conduct in-depth case studies of any particular research collaboration. Also, we can explore how homophily is established in interdisciplinary versus intradisciplinary research collaborations.

## References

- [1] A. A. Stoica, Homophily in co-authorship networks, *Int. Rev. Soc. Res.*, 8.2 (2018), 119–128.
- [2] H. Bisgin, N. Agarwal, X. Xu, A study of homophily on social media, *World Wide Web*, 15.2 (2012), 213–232.
- [3] K. Z. Khanam, G. Srivastava, V. Mago, The homophily principle in social network analysis, (2020), *arXiv preprint arXiv:2008.10383*.
- [4] N. E. D. Ferreyra, T. Hecking, E. Aïmeur, M. Heisel, H. U. Hoppe, Community Detection for Access-Control Decisions: Analysing the Role of Homophily and Information Diffusion in Online Social Networks, (2021) *arXiv preprint arXiv:2104.09137*.
- [5] H. Lu, Y. Feng, A measure of authors' centrality in co-authorship networks based on the distribution of collaborative relationships, *Scientometrics*, 81.2 (2009), 499–511.
- [6] M. G. Hâncean, M. Perc, Homophily in coauthorship networks of East European sociologists, *Scientific reports*, 6.1 (2016), 1–12.
- [7] M. Kwiek, W. Roszka, Gender-based homophily in research: A large-scale study of man-woman collaboration, *Journal of Informetrics*, 15.3 (2021), 101171.
- [8] K. Das, S. Samanta, M. Pal, Study on centrality measures in social networks: a survey, *Social network analysis and mining*, 8.1 (2018), 1–11.
- [9] F. Bloch, M. O. Jackson, P. Tebaldi, Centrality measures in networks, *arXiv preprint arXiv:1608.05845*, (2016).
- [10] M. McPherson, L. Smith-Lovin, J. M. Cook, Birds of a feather: Homophily in social networks, *Annual review of sociology*, (2001), 415–444.
- [11] B. S. Lawrence, N. P. Shah, Homophily: Measures and meaning. *Academy of Management Annals*, 14.2 (2020), 513–597.

- [12] R. Dwivedi, S. P. Nerur, Analyzing Co-authorship Network for Homophily-Evidence from IS senior Scholar's Basket of Eight Journals for Business Analytics Research. In *AMCIS*, (2020).
- [13] M. Gallivan, M. Ahuja, Co-authorship, homophily, and scholarly influence in information systems research. *Journal of the Association for Information Systems*, 16.12(2015), 2.
- [14] M. Cristani, D. Fogoroasi, C. Tomazzoli, Measuring Homophily. In *KDWeb*, (2016).
- [15] V. Umadevi, Case study–centrality measure analysis on co-authorship network. *Journal of Global Research in Computer Science*, 4.1 (2013), 67–70.
- [16] A. Dias, S. Ruthes, L. Lima, E. Campra, M. Silva, M. Bragança de Sousa, and G. Porto, Network centrality analysis in management and accounting sciences. *RAUSP Management Journal*, 55(2020), 207–226.
- [17] P. Hage, F. Harary, Eccentricity and centrality in networks. *Social networks*, 17.1 (1995), 57–63.
- [18] L. Holman, C. Morandin, Researchers collaborate with same-gendered colleagues more often than expected across the life sciences, *PloS one*, 14.4(2019), e0216128.
- [19] J. Fagan, K. S. Eddens, J. Dolly, N. L. Vanderford, H. Weiss, & J. S. Levens, Assessing research collaboration through co-authorship network analysis. *The journal of research administration*, 49.1(2018), 76.
- [20] M. H. Jones, T. S. Hackel, and R. A. Gross, The homophily and centrality of LGBTQ youth: A new story?, *Social Psychology of Education*, 25.5(2022), 1157–1175.
- [21] R. Mahapatra, S. Samanta, and M. Pal, Detecting influential node in a network using neutrosophic graph and its application. *Soft Computing*, 27(14), 9247–9260 (2023).
- [22] S. Samanta, V. K. Dubey, and B. Sarkar, Measure of influences in social networks. *Applied Soft Computing*, 99, 106858 (2021).
- [23] G. Muhiuddin, S. Samanta, A. F. Aljohani, and A. M. Alkhaibari, A study on graph centrality measures of different diseases due to DNA sequencing. *Mathematics*, 11(14), 3166 (2023).
- [24] S. D. Pandey, A. S. Ranadive, S. Samanta, & B. Sarkar, Bipolar-Valued Fuzzy Social Network and Centrality Measures. *Discrete Dynamics in Nature and Society*, 2022(1), 9713575.

## **Biographies**



**Laxminarayan Sahoo** is currently an Associate Professor of Computer and Information Science, at Raiganj University, Raiganj, India. He obtained his MSc from Vidyasagar University, India and his PhD from The University of Burdwan, India. He has received an MHRD fellowship from Govt. of India and received Prof. M.N. Gopalan Award for Best PhD thesis in Operations Research from the Operational Research Society of India (ORSI). He is a reviewer of several international journals and Academic Editor of the International Journal “Mathematical Problems in Engineering, Hindawi Publication. He is also an Associate Editor of “Journal of Graphic Era University” River Publication. His specializations include Wireless Sensor Networks, Distributed Computing, Reliability Optimization, Genetic Algorithms, Particle Swarm Optimization, Graph Theory, Fuzzy Game Theory, Interval Mathematics, Soft Computing, Fuzzy Decision making, and Operations Research. He has published a good number of articles in international and national journals of repute. Dr. Sahoo is the author of the books “Advanced Operations Research” published by Asian Books, New Delhi, “Advanced Optimization and Operations Research” published by Springer Nature, Singapore. He edited a book entitled Real Life Applications of Multiple Criteria Decision–Making Techniques in Fuzzy Domain” published by Springer Nature and wrote several chapters from reputed publishers like Springer, IGI Global, CRC Press, Walter de Gruyter, McGraw-Hill and Elsevier. He is a fellow of ISROSET.



**Sanchita Guchhait** received her M.C.A. from the University of Kalyani in 2011 and M.Phil. from Vidyasagar University in 2021. She is currently Pursuing a Ph.D. in Computer and Information Science from Raiganj University since 2022. She has more than 10 years of teaching experience.